7/5/2020

CATALOG

LEARNING VOYAGE

[CATEGORY Data Science 5](#_Toc51519814)

[Algorithms of the Intelligent Web 5](#_Toc51519815)

[Data Science Algorithms in a Week 7](#_Toc51519816)

[Jupyter for Data Science 9](#_Toc51519817)

[Machine Learning with R Cookbook 11](#_Toc51519818)

[Machine Learning Cookbook 15](#_Toc51519819)

[Beginning Data Science with Python and Jupyter 17](#_Toc51519820)

[Practical Data Science Cookbook 18](#_Toc51519821)

[Java for Data Science 21](#_Toc51519822)

[Python Social Media Analytics 23](#_Toc51519823)

[Learning Social Media Analytics with R 25](#_Toc51519824)

[Python for Finance 27](#_Toc51519825)

[Real-time Data Processing and Analytics 30](#_Toc51519826)

[Apache Spark 2.x Machine Learning Cookbook 32](#_Toc51519827)

[Python Machine Learning Cookbook 36](#_Toc51519828)

[TensorFlow Machine Learning Cookbook 40](#_Toc51519829)

[R Machine Learning Projects 42](#_Toc51519830)

[Large Scale Machine Learning with Python 44](#_Toc51519831)

[Exploratory Analysis with pandas 46](#_Toc51519832)

[Integrating Hadoop 48](#_Toc51519833)

[Machine Learning Solutions 50](#_Toc51519834)

[Turning Text into Gold: Taxonomies and Textual Analytics 53](#_Toc51519835)

[Keras 2.x Projects 55](#_Toc51519836)

[Ensemble Machine Learning Cookbook 57](#_Toc51519837)

[Intelligent Mobile Projects with TensorFlow 59](#_Toc51519838)

[Jupyter Cookbook 62](#_Toc51519839)

[Machine Learning for Algorithmic Trading 64](#_Toc51519840)

[Machine Learning with the Elastic Stack 67](#_Toc51519841)

[Python Machine Learning Blueprints 69](#_Toc51519842)

[Machine Learning for Mobile 71](#_Toc51519843)

[Data Analysis and Business Modeling with Excel 2013 73](#_Toc51519844)

[Build a Career in Data Science 75](#_Toc51519845)

[Data Science Boot camp 77](#_Toc51519846)

[Healthcare Analytics Made Simple 80](#_Toc51519847)

[Machine Learning for Healthcare Analytics Projects 82](#_Toc51519848)

[Machine Learning for Finance 84](#_Toc51519849)

[Go Machine Learning Projects 87](#_Toc51519850)

[Machine Learning Projects for Mobile Applications 89](#_Toc51519851)

[IoT Solutions with Blockchain 91](#_Toc51519852)

[Music Generation with Magenta 93](#_Toc51519853)

[Practical Recommender Systems 95](#_Toc51519854)

[Recommendation Systems with Python 98](#_Toc51519855)

[Natural Language Processing 100](#_Toc51519856)

[Relevant Search 102](#_Toc51519857)

[Think Like a Data Scientist 104](#_Toc51519858)

[Real-World Machine Learning. 106](#_Toc51519859)

[TensorFlow Machine Learning Projects 108](#_Toc51519860)

[Machine Learning to Detect Phishing Websites 111](#_Toc51519861)

[Monitoring Changes in Surface Water Using Satellite Image Data 113](#_Toc51519862)

[Human Pose Estimation with Deep Neural Networks 115](#_Toc51519863)

[Discovering Disease Outbreaks from News Headlines 117](#_Toc51519864)

[Decoding Data Science Job Postings to Improve Your Resume 119](#_Toc51519865)

[Practical Data Science with R 121](#_Toc51519866)

[Building Domain Specific Language Models 123](#_Toc51519867)

[3D Medical Image Analysis with PyTorch 125](#_Toc51519868)

[Growth Hacking with NLP and Sentiment Analysis 127](#_Toc51519869)

[Collective Intelligence 129](#_Toc51519870)

[Scikit-learn Cookbook 131](#_Toc51519871)

[Building Machine Learning Systems with Python 134](#_Toc51519872)

[SAS for Finance 136](#_Toc51519873)

[CATEGORY Programming 138](#_Toc51519874)

[R 138](#_Toc51519875)

[CATEGORY Artificial Intelligence 140](#_Toc51519876)

[Mahout 140](#_Toc51519877)

[Machine Learning on AWS 143](#_Toc51519878)

[Machine Learning for Developers 146](#_Toc51519879)

[Machine Learning with IBM Watson 148](#_Toc51519880)

[Scala for Machine Learning 150](#_Toc51519881)

[Deep-Learning-for-Natural-Language-Processing 153](#_Toc51519882)

[Ensemble Learning with R 155](#_Toc51519883)

[Machine Learning with scikit-learn 157](#_Toc51519884)

[Machine Learning for OpenCV 159](#_Toc51519885)

[Machine Learning for the Web 161](#_Toc51519886)

[Java Machine Learning 163](#_Toc51519887)

[Predictive Analytics with R 165](#_Toc51519888)

[Statistical Application Development with R and Python 167](#_Toc51519889)

[Learn Unity ML-Agents 169](#_Toc51519890)

[TensorFlow 2.0 171](#_Toc51519891)

[Deep Learning for Vision Systems 173](#_Toc51519892)

[Deep Learning with PyTorch 175](#_Toc51519893)

[Training Systems using Python Statistical Modeling 177](#_Toc51519894)

[Deep Reinforcement Learning 179](#_Toc51519895)

[Applied Unsupervised Learning with R 181](#_Toc51519896)

[Essential Natural Language Processing 183](#_Toc51519897)

[GANs 185](#_Toc51519898)

[Machine Learning with Apache Spark 187](#_Toc51519899)

[Machine Learning with Core ML 189](#_Toc51519900)

[Python Machine Learning By Example 191](#_Toc51519901)

[Bayesian Analysis with Python 194](#_Toc51519902)

[Machine Learning for Data Mining 196](#_Toc51519903)

[Machine Learning With Go 198](#_Toc51519904)

[Machine Learning with R 200](#_Toc51519905)

[Machine Learning with Scala 202](#_Toc51519906)

[Q-Learning with Python 204](#_Toc51519907)

[C# Machine Learning Projects 206](#_Toc51519908)

[Machine Learning with C# 208](#_Toc51519909)

[Machine Learning with Microsoft Excel 2019 210](#_Toc51519910)

[Machine Learning with JavaScript 212](#_Toc51519911)

[Ensemble Learning with Python 214](#_Toc51519912)

[Machine Learning Fundamentals 216](#_Toc51519913)

[Machine Learning in Java 218](#_Toc51519914)

[What's New in TensorFlow 2.0 220](#_Toc51519915)

[Blockchain Development for Finance Projects 222](#_Toc51519916)

[Machine Learning Bootcamp 225](#_Toc51519917)

[Machine Learning with scikit-learn 227](#_Toc51519918)

[Time Series Analysis with R 229](#_Toc51519919)

[Unsupervised Learning with Python 231](#_Toc51519920)

[Applied Unsupervised Learning with Python 233](#_Toc51519921)

[Machine Learning. 235](#_Toc51519922)

[Machine Learning Overview. 237](#_Toc51519923)

[Applied Supervised Learning with R. 239](#_Toc51519924)

[Machine Learning with TensorFlow.js. 242](#_Toc51519925)

[Python Machine Learning. 244](#_Toc51519926)

[Machine Learning Algorithms. 247](#_Toc51519927)

[One-shot Learning with Python 250](#_Toc51519928)

[Machine Learning with Go 252](#_Toc51519929)

[Applied SQL Data Analytics 254](#_Toc51519930)

[Supervised Learning 256](#_Toc51519931)

[Supervised Machine Learning with Python 258](#_Toc51519932)

[Machine Learning with C++. 260](#_Toc51519933)

[Natural Language Processing for Hackers 263](#_Toc51519934)

[Machine Learning with ML.NET. 265](#_Toc51519935)

[Machine Learning with TensorFlow 267](#_Toc51519936)

[Zero to AI 269](#_Toc51519937)

[Deep Learning with R 271](#_Toc51519938)

[Deep Learning with Python 273](#_Toc51519939)

[Machine Learning. 275](#_Toc51519940)

[CATEGORY Data Analysis / BI 278](#_Toc51519941)

[Tika 278](#_Toc51519942)

[Pandas 279](#_Toc51519943)

[Learning pandas 282](#_Toc51519944)

[Data Analysis with NumPy and Pandas 285](#_Toc51519945)

[Turning Spreadsheets into Corporate Data 287](#_Toc51519946)

[Learning Jupyter 5 289](#_Toc51519947)

[Data Wrangling with Python 291](#_Toc51519948)

[Practical Data Wrangling 293](#_Toc51519949)

[Python Data Structures and Algorithms 295](#_Toc51519950)

[Data Analysis with Python 297](#_Toc51519951)

[Go Web Scraping 299](#_Toc51519952)

[Feature Engineering Made Easy 301](#_Toc51519953)

[Data Analysis with Scala 303](#_Toc51519954)

[Learning Alteryx 305](#_Toc51519955)

[Haskell Data Analysis 307](#_Toc51519956)

[Python Feature Engineering Cookbook 309](#_Toc51519957)

[SciPy Recipes 312](#_Toc51519958)

[Exploratory Data Analysis with Python 315](#_Toc51519959)

[SAS for Data Analysis 317](#_Toc51519960)

[R Data Analysis Projects 319](#_Toc51519961)

[Data Analysis with Pandas 321](#_Toc51519962)

[Exploratory Data Analysis with R 323](#_Toc51519963)

[Data Analysis with R 325](#_Toc51519964)

[Exploring the Data Jungle 328](#_Toc51519965)

[Visualizing Graph Data 329](#_Toc51519966)

[Taming Text 331](#_Toc51519967)

[D3.js 333](#_Toc51519968)

[Web Scraping with Python 335](#_Toc51519969)

[CATEGORY Big Data 337](#_Toc51519970)

[Redis. 337](#_Toc51519971)

[Machine Learning with Spark 2.x 338](#_Toc51519972)

[Solr. 341](#_Toc51519973)

[Hadoop in Practice 343](#_Toc51519974)

[Apache Hadoop 3 345](#_Toc51519975)

[Apache Hive Essentials 347](#_Toc51519976)

[Storm Applied. 349](#_Toc51519977)

[Big Data 351](#_Toc51519978)

[Apache Spark 353](#_Toc51519979)

[Building Data Streaming Applications with Apache Kafka 355](#_Toc51519980)

[Spark GraphX. 358](#_Toc51519981)

[Streaming Data 360](#_Toc51519982)

[CATEGORY DevOps 362](#_Toc51519983)

[Elasticsearch 362](#_Toc51519984)

[AI as a Service 364](#_Toc51519985)

[Introduction to DevOps with Kubernetes 366](#_Toc51519986)

[Automated Machine Learning 368](#_Toc51519987)

[Real-World Cryptography 370](#_Toc51519988)

[CATEGORY Cloud 372](#_Toc51519989)

[Machine Learning with Azure 372](#_Toc51519990)

[Cloud Analytics with Microsoft Azure 374](#_Toc51519991)

[Machine Learning on Google Cloud Platform 376](#_Toc51519992)

[Machine Learning with AWS 378](#_Toc51519993)

[CATEGORY CyberSecurity 380](#_Toc51519994)

[Machine Learning for Cybersecurity 380](#_Toc51519995)

[Applied Dark Web Analysis 382](#_Toc51519996)



CATEGORY Data Science

# Algorithms of the Intelligent Web

Course Snapshot

* **Course:** Algorithms of the Intelligent Web
* **Duration:** 2 days
* **Skill-level**: Foundation-level Algorithms of the Intelligent Web skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to know how to create machine learning applications that crunch and wrangle data collected from users, web applications, and website logs.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Algorithms of the Intelligent Web** teaches you how to create machine learning applications that crunch and wrangle data collected from users, web applications, and website logs. In this totally revised edition, you’ll look at intelligent algorithms that extract real value from data. Key machine learning concepts are explained with code examples in Pythons scikit-learn. This course guides you through algorithms to capture, store, and structure data streams coming from the web. You’ll explore recommendation engines and dive into classification via statistical algorithms, neural networks, and deep learning.

Working in a hands-on learning environment, led by our Algorithms of the Intelligent Web expert instructor, students will learn about and explore:

* teaches the most important approaches to algorithmic web data analysis, enabling you to create your own machine
* learning applications that crunch, munge, and wrangle data collected from users, web applications, sensors and website logs.
* how to create machine learning applications that crunch and wrangle data collected from users, web applications, and website logs

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Building applications for the intelligent web
* Extracting structure from data: clustering and transforming your data
* Recommending relevant content
* Classification: placing things where they belong
* Case study: click prediction for online advertising
* Deep learning and neural networks
* Making the right choice
* The future of the intelligent web

Audience & Pre-Requisites

This course is designed for developers interested in data science and for beginner data scientists

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

2. **Building applications for the intelligent web**

An intelligent algorithm: Google Now

* The intelligent-algorithm lifecycle
* Further examples of intelligent algorithms
* Things that intelligent applications are not
* Classes of intelligent algorithm
* Evaluating the performance of intelligent algorithms
* Important notes about intelligent algorithms

1. **Extracting structure from data: clustering and transforming your data**

* Data, structure, bias, and noise
* The curse of dimensionality
* K-means
* The Gaussian mixture models
* The relationship between k-means and GMM
* Transforming the data axis

1. **Recommending relevant content**

* Setting the scene: an online movie store
* Distance and similarity
* How do recommendation engines work?
* User-based collaborative filtering
* Model-based recommendation using singular value decomposition
* The Netflix Prize
* Evaluating your recommender

1. **Classification: placing things where they belong**

* The need for classification
* An overview of classifiers
* Fraud detection with logistic regression
* Are your results credible?
* Classification with very large datasets

1. **Case study: click prediction for online advertising**

* History and background
* The exchange
* What is a bidder?
* What is a decisioning engine?
* Click prediction with Vowpal Wabbit
* Complexities of building a decisioning engine
* The future of real-time prediction

1. **Deep learning and neural networks**

* An intuitive approach to deep learning
* Neural networks
* The perceptron
* Multilayer perceptrons
* Going deeper: from multilayer neural networks to deep learning

1. **Making the right choice**

* A/B testing
* Multi-armed bandits
* Bayesian bandits in the wild
* A/B vs. the Bayesian bandit
* Extensions to multi-armed bandits

1. **The future of the intelligent web**

* Future applications of the intelligent web
* Social implications of the intelligent web

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Science Algorithms in a Week

Course Snapshot

* **Course:** Data Science Algorithms in a Week
* **Duration:** 3 days
* **Skill-level**: Foundation-level Data Science Algorithms skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Build strong foundation of machine learning algorithms In 7 days.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning applications are highly automated and self-modifying, and they continue to improve over time with minimal human intervention as they learn with more data. To address the complex nature of various real-world data problems, specialized machine learning algorithms have been developed that solve these problems perfectly. Data science helps you gain new knowledge from existing data through algorithmic and statistical analysis. This course will address the problems related to accurate and efficient data classification and prediction. Over the course of 7 days, you will be introduced to seven algorithms, along with exercises that will help you learn different aspects of machine learning. You will see how to pre-cluster your data to optimize and classify it for large datasets. You will then find out how to predict data based on the existing trends in your datasets. This course covers algorithms such as: k-Nearest Neighbors, Naive Bayes, Decision Trees, Random Forest, k-Means, Regression, and Time-series. On completion of the book, you will understand which machine learning algorithm to pick for clustering, classification, or regression and which is best suited for your problem.

Working in a hands-on learning environment, led by our Data Science Algorithms expert instructor, students will learn about and explore:

* Get to know seven algorithms for your data science needs in this concise, insightful guide
* Ensure you’re confident in the basics by learning when and where to use various data science algorithms
* Learn to use machine learning algorithms in a period of just 7 days

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Find out how to classify using Naive Bayes, Decision Trees, and Random Forest to achieve accuracy to solve complex problems
* Identify a data science problem correctly and devise an appropriate prediction solution using Regression and Time-series
* See how to cluster data using the k-Means algorithm
* Get to know how to implement the algorithms efficiently in the Python and R languages

Audience & Pre-Requisites

This course is designed for developers interested to Build strong foundation of machine learning algorithms In 7 days.

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Classification Using K Nearest Neighbors**

* Classification Using K Nearest Neighbors
* Mary and her temperature preferences
* Implementation of k-nearest neighbors algorithm
* Map of Italy example - choosing the value of k
* House ownership - data rescaling
* Text classification - using non-Euclidean distances
* Text classification - k-NN in higher-dimensions

1. **Naive Bayes**

* Naive Bayes
* Medical test - basic application of Bayes' theorem
* Proof of Bayes' theorem and its extension
* Playing chess - independent events
* Implementation of naive Bayes classifier
* Playing chess - dependent events
* Gender classification - Bayes for continuous random variables

1. **Decision Trees**

* Decision Trees
* Swim preference - representing data with decision tree
* Information theory
* ID3 algorithm - decision tree construction
* Classifying with a decision tree
* Playing chess - analysis with decision tree
* Going shopping - dealing with data inconsistency

1. **Random Forest**

* Random Forest
* Overview of random forest algorithm
* Swim preference - analysis with random forest
* Implementation of random forest algorithm
* Playing chess example
* Going shopping - overcoming data inconsistency with randomness and measuring the level of confidence

1. **Clustering into K Clusters**

* Clustering into K Clusters
* Household incomes - clustering into k clusters
* Gender classification - clustering to classify
* Implementation of the k-means clustering algorithm
* House ownership – choosing the number of clusters
* Document clustering – understanding the number of clusters k in a semantic context

1. **Regression**

* Regression
* Fahrenheit and Celsius conversion - linear regression on perfect data
* Weight prediction from height - linear regression on real-world data
* Gradient descent algorithm and its implementation
* Flight time duration prediction from distance
* Ballistic flight analysis – non-linear model

1. **Time Series Analysis**

* Time Series Analysis
* Business profit - analysis of the trend
* Electronics shop's sales - analysis of seasonality

1. **Statistics**

* Statistics
* Basic concepts
* Bayesian Inference
* Distributions
* Cross-validation
* A/B Testing

1. **R Reference**

* R Reference
* Introduction
* Data types
* Linear regression

1. **Python Reference**

* Python Reference
* Introduction
* Data types
* Flow control

1. **Glossary of Algorithms and Methods in Data Science**

* Glossary of Algorithms and Methods in Data Science

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Jupyter for Data Science

Course Snapshot

* **Course:** Jupyter for Data Science
* **Duration:** 3 days
* **Skill-level**: Foundation-level Jupyter for Data Science skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Get guide to building an efficient data science pipeline using Jupyter.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Jupyter Notebook is a web-based environment that enables interactive computing in notebook documents. It allows you to create documents that contain live code, equations, and visualizations. This book is a comprehensive guide to getting started with data science using the popular Jupyter notebook. If you are familiar with Jupyter notebook and want to learn how to use its capabilities to perform various data science tasks, this is the book for you! From data exploration to visualization, this book will take you through every step of the way in implementing an effective data science pipeline using Jupyter. You will also see how you can utilize Jupyter's features to share your documents and codes with your colleagues. The book also explains how Python 3, R, and Julia can be integrated with Jupyter for various data science tasks. By the end of this book, you will comfortably leverage the power of Jupyter to perform various tasks in data science successfully.

Working in a hands-on learning environment, led by our Jupyter for Data Science expert instructor, students will learn about and explore:

* Get the most out of your Jupyter notebook to complete the trickiest of tasks in Data Science
* Learn all the tasks in the data science pipeline—from data acquisition to visualization—and implement them using Jupyter
* Get ahead of the curve by mastering all the applications of Jupyter for data science with this unique and intuitive guide

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand why Jupyter notebooks are a perfect fit for your data science tasks
* Perform scientific computing and data analysis tasks with Jupyter
* Interpret and explore different kinds of data visually with charts, histograms, and more
* Extend SQL's capabilities with Jupyter notebooks
* Combine the power of R and Python 3 with Jupyter to create dynamic notebooks
* Create interactive dashboards and dynamic presentations
* Master the best coding practices and deploy your Jupyter notebooks efficiently

Audience & Pre-Requisites

This course is designed for beginners who wants to guide to building an efficient data science pipeline using Jupyter.

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Jupyter and Data Science**

* Jupyter and Data Science
* Jupyter concepts
* A first look at the Jupyter user interface

1. **Working with Analytical Data on Jupyter**

* Working with Analytical Data on Jupyter
* Data scraping with a Python notebook
* Using heavy-duty data processing functions in Jupyter
* Using SciPy in Jupyter
* Expanding on panda data frames in Jupyter

1. **Data Visualization and Prediction**

* Data Visualization and Prediction
* Make a prediction using scikit-learn
* Make a prediction using R
* Interactive visualization
* Plotting using Plotly
* Creating a human density map
* Draw a histogram of social data
* Plotting 3D data

1. **Data Mining and SQL Queries**

* Data Mining and SQL Queries
* Special note for Windows installation
* Using Spark to analyze data
* Another MapReduce example
* Using SparkSession and SQL
* Combining datasets
* Loading JSON into Spark
* Using Spark pivot

1. **R with Jupyter**

* R with Jupyter
* How to set up R for Jupyter
* R data analysis of the 2016 US election demographics
* Analyzing 2016 voter registration and voting
* Analyzing changes in college admissions
* Predicting airplane arrival time

1. **Data Wrangling**

* Data Wrangling
* Reading a CSV file
* Reading another CSV file
* Manipulating data with dplyr
* Sampling a dataset
* Tidying up data with tidyr

1. **Jupyter Dashboards**

* Jupyter Dashboards
* Visualizing glyph ready data
* Publishing a notebook
* Creating a Shiny dashboard
* Building standalone dashboards

1. **Statistical Modeling**

* Statistical Modeling
* Converting JSON to CSV
* Evaluating Yelp reviews
* Using Python to compare ratings
* Visualizing average ratings by cuisine
* Arbitrary search of ratings
* Determining relationships between number of ratings and ratings

1. **Machine Learning Using Jupyter**

* Machine Learning Using Jupyter
* Naive Bayes
* Nearest neighbor estimator
* Decision trees
* Neural networks
* Random forests

1. **Optimizing Jupyter Notebooks**

* Optimizing Jupyter Notebooks
* Deploying notebooks
* Optimizing your script
* Monitoring Jupyter
* Caching your notebook
* Securing a notebook
* Scaling Jupyter Notebooks
* Sharing Jupyter Notebooks
* Converting a notebook
* Versioning a notebook

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with R Cookbook

Course Snapshot

* **Course:** Machine Learning with R Cookbook
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning with R Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Explore over 110 recipes to analyze data and build predictive models with simple and easy-to-use R code.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Big data has become a popular buzzword across many industries. An increasing number of people have been exposed to the term and are looking at how to leverage big data in their own businesses, to improve sales and profitability. However, collecting, aggregating, and visualizing data is just one part of the equation. Being able to extract useful information from data is another task, and a much more challenging one. Machine Learning with R Cookbook, Second Edition uses a practical approach to teach you how to perform machine learning with R. Each chapter is divided into several simple recipes. Through the step-by-step instructions provided in each recipe, you will be able to construct a predictive model by using a variety of machine learning packages. In this book, you will first learn to set up the R environment and use simple R commands to explore data. The next topic covers how to perform statistical analysis with machine learning analysis and assess created models, covered in detail later on in the book. You'll also learn how to integrate R and Hadoop to create a big data analysis platform. The detailed illustrations provide all the information required to start applying machine learning to individual projects. With Machine Learning with R Cookbook, machine learning has never been easier.

Working in a hands-on learning environment, led by our ML expert instructor, students will learn about and explore:

* Apply R to simplify predictive modeling with short and simple code
* Use machine learning to solve problems ranging from small to big data
* Build a training and testing dataset, applying different classification methods.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Create and inspect transaction datasets and perform association analysis with the Apriori algorithm
* Visualize patterns and associations using a range of graphs and find frequent item-sets using the Eclat algorithm
* Compare differences between each regression method to discover how they solve problems
* Detect and impute missing values in air quality data
* Predict possible churn users with the classification approach
* Plot the autocorrelation function with time series analysis
* Use the Cox proportional hazards model for survival analysis
* Implement the clustering method to segment customer data
* Compress images with the dimension reduction method
* Incorporate R and Hadoop to solve machine learning problems on big data

Audience & Pre-Requisites

This course is designed for developers interested to Explore over 110 recipes to analyze data and build predictive models with simple and easy-to-use R code

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Practical Machine Learning with R**

* Practical Machine Learning with R
* Introduction
* Downloading and installing R
* Downloading and installing RStudio
* Installing and loading packages
* Understanding of basic data structures
* Basic commands for subsetting
* Reading and writing data
* Manipulating data
* Applying basic statistics
* Visualizing data
* Getting a dataset for machine learning

1. **Data Exploration with Air Quality Datasets**

* Data Exploration with Air Quality Datasets
* Introduction
* Using air quality dataset
* Converting attributes to factor
* Detecting missing values
* Imputing missing values
* Exploring and visualizing data
* Predicting values from datasets

1. **Analyzing Time Series Data**

* Analyzing Time Series Data
* Introduction
* Looking at time series data
* Plotting and forecasting time series data
* Extracting, subsetting, merging, filling, and padding
* Successive differences and moving averages
* Exponential smoothing
* Plotting the autocorrelation function

1. **R and Statistics**

* R and Statistics
* Introduction
* Understanding data sampling in R
* Operating a probability distribution in R
* Working with univariate descriptive statistics in R
* Performing correlations and multivariate analysis
* Conducting an exact binomial test
* Performing a student's t-test
* Performing the Kolmogorov-Smirnov test
* Understanding the Wilcoxon Rank Sum and Signed Rank test
* Working with Pearson's Chi-squared test
* Conducting a one-way ANOVA
* Performing a two-way ANOVA

1. **Understanding Regression Analysis**

* Understanding Regression Analysis
* Introduction
* Different types of regression
* Fitting a linear regression model with lm
* Summarizing linear model fits
* Using linear regression to predict unknown values
* Generating a diagnostic plot of a fitted model
* Fitting multiple regression
* Summarizing multiple regression
* Using multiple regression to predict unknown values
* Fitting a polynomial regression model with lm
* Fitting a robust linear regression model with rlm
* Studying a case of linear regression on SLID data
* Applying the Gaussian model for generalized linear regression
* Applying the Poisson model for generalized linear regression
* Applying the Binomial model for generalized linear regression
* Fitting a generalized additive model to data
* Visualizing a generalized additive model
* Diagnosing a generalized additive model

1. **Survival Analysis**

* Survival Analysis
* Introduction
* Loading and observing data
* Viewing the summary of survival analysis
* Visualizing the Survival Curve
* Using the log-rank test
* Using the COX proportional hazard model
* Nelson-Aalen Estimator of cumulative hazard

1. **Classification 1 - Tree, Lazy, and Probabilistic**

* Classification 1 - Tree, Lazy, and Probabilistic
* Introduction
* Preparing the training and testing datasets
* Building a classification model with recursive partitioning trees
* Visualizing a recursive partitioning tree
* Measuring the prediction performance of a recursive partitioning tree
* Pruning a recursive partitioning tree
* Handling missing data and split and surrogate variables
* Building a classification model with a conditional inference tree
* Control parameters in conditional inference trees
* Visualizing a conditional inference tree
* Measuring the prediction performance of a conditional inference tree
* Classifying data with the k-nearest neighbor classifier
* Classifying data with logistic regression
* Classifying data with the Naïve Bayes classifier

1. **Classification 2 - Neural Network and SVM**

* Classification 2 - Neural Network and SVM
* Introduction
* Classifying data with a support vector machine
* Choosing the cost of a support vector machine
* Visualizing an SVM fit
* Predicting labels based on a model trained by a support vector machine
* Tuning a support vector machine
* The basics of neural network
* Training a neural network with neuralnet
* Visualizing a neural network trained by neuralnet
* Predicting labels based on a model trained by neuralnet
* Training a neural network with nnet
* Predicting labels based on a model trained by nnet

1. **Model Evaluation**

* Model Evaluation
* Introduction
* Estimating model performance with k-fold cross-validation
* Estimating model performance with Leave One Out Cross Validation
* Performing cross-validation with the e1071 package
* Performing cross-validation with the caret package
* Ranking the variable importance with the caret package
* Ranking the variable importance with the rminer package
* Finding highly correlated features with the caret package
* Selecting features using the caret package
* Measuring the performance of the regression model
* Measuring prediction performance with a confusion matrix
* Measuring prediction performance using ROCR
* Comparing an ROC curve using the caret package
* Measuring performance differences between models with the caret package

1. **Ensemble Learning**

* Ensemble Learning
* Introduction
* Using the Super Learner algorithm
* Using ensemble to train and test
* Classifying data with the bagging method
* Performing cross-validation with the bagging method
* Classifying data with the boosting method
* Performing cross-validation with the boosting method
* Classifying data with gradient boosting
* Calculating the margins of a classifier
* Calculating the error evolution of the ensemble method
* Classifying data with random forest
* Estimating the prediction errors of different classifiers

1. **Clustering**

* Clustering
* Introduction
* Clustering data with hierarchical clustering
* Cutting trees into clusters
* Clustering data with the k-means method
* Drawing a bivariate cluster plot
* Comparing clustering methods
* Extracting silhouette information from clustering
* Obtaining the optimum number of clusters for k-means
* Clustering data with the density-based method
* Clustering data with the model-based method
* Visualizing a dissimilarity matrix
* Validating clusters externally

1. **Association Analysis and Sequence Mining**

* Association Analysis and Sequence Mining
* Introduction
* Transforming data into transactions
* Displaying transactions and associations
* Mining associations with the Apriori rule
* Pruning redundant rules
* Visualizing association rules
* Mining frequent itemsets with Eclat
* Creating transactions with temporal information
* Mining frequent sequential patterns with cSPADE
* Using the TraMineR package for sequence analysis
* Visualizing sequence, Chronogram, and Traversal Statistics

1. **Dimension Reduction**

* Dimension Reduction
* Introduction
* Why to reduce the dimension?
* Performing feature selection with FSelector
* Performing dimension reduction with PCA
* Determining the number of principal components using the scree test
* Determining the number of principal components using the Kaiser method
* Visualizing multivariate data using biplot
* Performing dimension reduction with MDS
* Reducing dimensions with SVD
* Compressing images with SVD
* Performing nonlinear dimension reduction with ISOMAP
* Performing nonlinear dimension reduction with Local Linear Embedding

1. **Big Data Analysis (R and Hadoop)**

* Big Data Analysis (R and Hadoop)
* Introduction
* Preparing the RHadoop environment
* Installing rmr2
* Installing rhdfs
* Operating HDFS with rhdfs
* Implementing a word count problem with RHadoop
* Comparing the performance between an R MapReduce program and a standard R program
* Testing and debugging the rmr2 program
* Installing plyrmr
* Manipulating data with plyrmr
* Conducting machine learning with RHadoop
* Configuring RHadoop clusters on Amazon EMR

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning Cookbook

Course Snapshot

* **Course:** Machine Learning Cookbook
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Build Machine Learning applications with R.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning has become the new black. The challenge in today’s world is the explosion of data from existing legacy data and incoming new structured and unstructured data. The complexity of discovering, understanding, performing analysis, and predicting outcomes on the data using machine learning algorithms is a challenge. This cookbook will help solve everyday challenges you face as a data scientist. The application of various data science techniques and on multiple data sets based on real-world challenges you face will help you appreciate a variety of techniques used in various situations. The first half of the course provides recipes on fairly complex machine-learning systems, where you’ll learn to explore new areas of applications of machine learning and improve its efficiency. That includes recipes on classifications, neural networks, unsupervised and supervised learning, deep learning, reinforcement learning, and more. The second half of the course focuses on three different machine learning case studies, all based on real-world data, and offers solutions and solves specific machine-learning issues in each one.

Working in a hands-on learning environment, led by our Machine Learning Cookbook expert instructor, students will learn about and explore:

* Implement a wide range of algorithms and techniques for tackling complex data
* Improve predictions and recommendations to have better levels of accuracy
* Optimize performance of your machine-learning systems

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get equipped with a deeper understanding of how to apply machine-learning techniques
* Implement each of the advanced machine-learning techniques
* Solve real-life problems that are encountered in order to make your applications produce improved results
* Gain hands-on experience in problem solving for your machine-learning systems
* Understand the methods of collecting data, preparing data for usage, training the model, evaluating the model’s performance, and improving the model’s performance

Audience & Pre-Requisites

This course is designed for developers interested in Building Machine Learning applications with R

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Introduction to Machine Learning**

* Introduction to Machine Learning
* What is machine learning?
* An overview of classification
* An overview of clustering
* An overview of supervised learning
* An overview of unsupervised learning
* An overview of reinforcement learning
* An overview of structured prediction
* An overview of neural networks
* An overview of deep learning

1. **Classification**

* Classification
* Introduction
* Discriminant function analysis - geological measurements on brines from wells
* Multinomial logistic regression - understanding program choices made by students
* Tobit regression - measuring the students' academic aptitude
* Poisson regression - understanding species present in Galapagos Islands

1. **Clustering**

* Clustering
* Introduction
* Hierarchical clustering - World Bank sample dataset
* Hierarchical clustering - Amazon rainforest burned between 1999-2010
* Hierarchical clustering - gene clustering
* Binary clustering - math test
* K-means clustering - European countries protein consumption
* K-means clustering – foodstuff

1. **Model Selection and Regularization**

* Model Selection and Regularization
* Introduction
* Shrinkage methods - calories burned per day
* Dimension reduction methods - Delta's Aircraft Fleet
* Principal component analysis - understanding world cuisine

1. **Nonlinearity**

* Nonlinearity
* Generalized additive models - measuring the household income of New Zealand
* Smoothing splines - understanding cars and speed
* Local regression - understanding drought warnings and impact

1. **Supervised Learning**

* Supervised Learning
* Introduction
* Decision tree learning - Advance Health Directive for patients with chest pain
* Decision tree learning - income-based distribution of real estate values
* Decision tree learning - predicting the direction of stock movement
* Naive Bayes - predicting the direction of stock movement
* Random forest - currency trading strategy
* Support vector machine - currency trading strategy
* Stochastic gradient descent - adult income

1. **Unsupervised Learning**

* Unsupervised Learning
* Introduction
* Self-organizing map - visualizing of heatmaps
* Vector quantization - image clustering

1. **Reinforcement Learning**

* Reinforcement Learning
* Introduction
* Markov chains - the stocks regime switching model
* Markov chains - the multi-channel attribution model
* Markov chains - the car rental agency service
* Continuous Markov chains - vehicle service at a gas station
* Monte Carlo simulations - calibrated Hull and White short-rates

1. **Structured Prediction**

* Structured Prediction
* Introduction
* Hidden Markov models - EUR and USD
* Hidden Markov models - regime detection

1. **Neural Networks**

* Neural Networks
* Introduction
* Modelling SP 500
* Measuring the unemployment rate

1. **Deep Learning**

* Deep Learning
* Introduction
* Recurrent neural networks - predicting periodic signals

1. **Case Study - Exploring World Bank Data**

* Case Study - Exploring World Bank Data
* Introduction
* Exploring World Bank data

1. **Case Study - Pricing Reinsurance Contracts**

* Case Study - Pricing Reinsurance Contracts
* Introduction
* Pricing reinsurance contracts

1. **Case Study - Forecast of Electricity Consumption**

* Case Study - Forecast of Electricity Consumption
* Introduction

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Beginning Data Science with Python and Jupyter

Course Snapshot

* **Course:** Beginning Data Science with Python and Jupyter
* **Duration:** 2 days
* **Skill-level**: Foundation-level Data Science with Python and Jupyter skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Get started with data science doesn't have to be an uphill battle.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Get to grips with the skills you need for entry-level data science in this hands-on Python and Jupyter course. You'll learn about some of the most commonly used libraries that are part of the Anaconda distribution, and then explore machine learning models with real datasets to give you the skills and exposure you need for the real world. We'll finish up by showing you how easy it can be to scrape and gather your own data from the open web, so that you can apply your new skills in an actionable context.

Working in a hands-on learning environment, led by our Data Science with Python and Jupyter expert instructor, students will learn about and explore:

* Get up and running with the Jupyter ecosystem and some example datasets
* Learn about key machine learning concepts like SVM, KNN classifiers and Random Forests
* Discover how you can use web scraping to gather and parse your own bespoke datasets

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Identify potential areas of investigation and perform exploratory data analysis
* Plan a machine learning classification strategy and train classification models
* Use validation curves and dimensionality reduction to tune and enhance your models
* Scrape tabular data from web pages and transform it into Pandas DataFrames
* Create interactive, web-friendly visualizations to clearly communicate your findings

Audience & Pre-Requisites

This course is designed for for beginners who know a little Python and are looking for a quick, fast-paced introduction.

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Jupyter Fundamentals**

* Jupyter Fundamentals
* Lesson Objectives
* Basic Functionality and Features
* Our First Analysis - The Boston Housing Dataset

1. **Data Cleaning and Advanced Machine Learning**

* Data Cleaning and Advanced Machine Learning
* Preparing to Train a Predictive Model
* Training Classification Models

1. **Web Scraping and Interactive Visualizations**

* Web Scraping and Interactive Visualizations
* Lesson Objectives
* Scraping Web Page Data
* Interactive Visualizations

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Practical Data Science Cookbook

Course Snapshot

* **Course:** Practical Data Science Cookbook
* **Duration:** 3 days
* **Skill-level**: Foundation-level Practical Data Science Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to complete real-world data science projects in R and Python
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

As increasing amounts of data are generated each year, the need to analyze and create value out of it is more important than ever. Companies that know what to do with their data and how to do it well will have a competitive advantage over companies that don’t. Because of this, there will be an increasing demand for people that possess both the analytical and technical abilities to extract valuable insights from data and create valuable solutions that put those insights to use. Starting with the basics, this course covers how to set up your numerical programming environment, introduces you to the data science pipeline, and guides you through several data projects in a step-by-step format. By sequentially working through the steps in each lesson, you will quickly familiarize yourself with the process and learn how to apply it to a variety of situations with examples using the two most popular programming languages for data analysis—R and Python.

Working in a hands-on learning environment, led by our Practical Data Science Cookbook expert instructor, students will learn about and explore:

* Tackle every step in the data science pipeline and use it to acquire, clean, analyze, and visualize your data
* Get beyond the theory and implement real-world projects in data science using R and Python
* Easy-to-follow recipes will help you understand and implement the numerical computing concepts

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn and understand the installation procedure and environment required for R and Python on various platforms
* Prepare data for analysis by implement various data science concepts such as acquisition, cleaning and munging through R and Python
* Build a predictive model and an exploratory model
* Analyze the results of your model and create reports on the acquired data
* Build various tree-based methods and Build random forest

Audience & Pre-Requisites

This course is designed for beginners who wants to complete real-world data science projects in R and Python

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Preparing Your Data Science Environment**

* Preparing Your Data Science Environment
* Understanding the data science pipeline
* Installing R on Windows, Mac OS X, and Linux
* Installing libraries in R and RStudio
* Installing Python on Linux and Mac OS X
* Installing Python on Windows
* Installing the Python data stack on Mac OS X and Linux
* Installing extra Python packages
* Installing and using virtualenv

1. **Driving Visual Analysis with Automobile Data with R**

* Driving Visual Analysis with Automobile Data with R
* Introduction
* Acquiring automobile fuel efficiency data
* Preparing R for your first project
* Importing automobile fuel efficiency data into R
* Exploring and describing fuel efficiency data
* Analyzing automobile fuel efficiency over time
* Investigating the makes and models of automobiles

1. **Creating Application-Oriented Analyses Using Tax Data and Python**

* Creating Application-Oriented Analyses Using Tax Data and Python
* Introduction
* Preparing for the analysis of top incomes
* Importing and exploring the world's top incomes dataset
* Analyzing and visualizing the top income data of the US
* Furthering the analysis of the top income groups of the US
* Reporting with Jinja2
* Repeating the analysis in R

1. **Modeling Stock Market Data**

* Modeling Stock Market Data
* Introduction
* Acquiring stock market data
* Summarizing the data
* Cleaning and exploring the data
* Generating relative valuations
* Screening stocks and analyzing historical prices

1. **Visually Exploring Employment Data**

* Visually Exploring Employment Data
* Introduction
* Preparing for analysis
* Importing employment data into R
* Exploring the employment data
* Obtaining and merging additional data
* Adding geographical information
* Extracting state- and county-level wage and employment information
* Visualizing geographical distributions of pay
* Exploring where the jobs are, by industry
* Animating maps for a geospatial time series
* Benchmarking performance for some common tasks

1. **Driving Visual Analyses with Automobile Data**

* Driving Visual Analyses with Automobile Data
* Introduction
* Getting started with IPython
* Exploring Jupyter Notebook
* Preparing to analyze automobile fuel efficiencies
* Exploring and describing fuel efficiency data with Python
* Analyzing automobile fuel efficiency over time with Python
* Investigating the makes and models of automobiles with Python

1. **Working with Social Graphs**

* Working with Social Graphs
* Introduction
* Preparing to work with social networks in Python
* Importing networks
* Exploring subgraphs within a heroic network
* Finding strong ties
* Finding key players
* Exploring the characteristics of entire networks
* Clustering and community detection in social networks
* Visualizing graphs
* Social networks in R

1. **Recommending Movies at Scale (Python)**

* Recommending Movies at Scale (Python)
* Introduction
* Modeling preference expressions
* Understanding the data
* Ingesting the movie review data
* Finding the highest-scoring movies
* Improving the movie-rating system
* Measuring the distance between users in the preference space
* Computing the correlation between users
* Finding the best critic for a user
* Predicting movie ratings for users
* Collaboratively filtering item by item
* Building a non-negative matrix factorization model
* Loading the entire dataset into the memory
* Dumping the SVD-based model to the disk
* Training the SVD-based model
* Testing the SVD-based model

1. **Harvesting and Geolocating Twitter Data (Python)**

* Harvesting and Geolocating Twitter Data (Python)
* Introduction
* Creating a Twitter application
* Understanding the Twitter API v1.1
* Determining your Twitter followers and friends
* Pulling Twitter user profiles
* Making requests without running afoul of Twitter's rate limits
* Storing JSON data to disk
* Setting up MongoDB for storing Twitter data
* Storing user profiles in MongoDB using PyMongo
* Exploring the geographic information available in profiles
* Plotting geospatial data in Python

1. **Forecasting New Zealand Overseas Visitors**

* Forecasting New Zealand Overseas Visitors
* Introduction
* The ts object
* Visualizing time series data
* Simple linear regression models
* ACF and PACF
* ARIMA models
* Accuracy measurements
* Fitting seasonal ARIMA models

1. **German Credit Data Analysis**

* German Credit Data Analysis
* Introduction
* Simple data transformations
* Visualizing categorical data
* Discriminant analysis
* Dividing the data and the ROC
* Fitting the logistic regression model
* Decision trees and rules
* Decision tree for german data

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Java for Data Science

Course Snapshot

* **Course:** Java for Data Science
* **Duration:** 3 days
* **Skill-level**: Foundation-level Java for Data Science and Jupyter skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to use Java to create a diverse range of Data Science applications and bring Data Science into production.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Java is the most popular programming language, according to the TIOBE index, and it is a typical choice for running production systems in many companies, both in the startup world and among large enterprises. Not surprisingly, it is also a common choice for creating data science applications: it is fast and has a great set of data processing tools, both built-in and external. What is more, choosing Java for data science allows you to easily integrate solutions with existing software, and bring data science into production with less effort. This course will teach you how to create data science applications with Java. First, we will revise the most important things when starting a data science application, and then brush up the basics of Java and machine learning before diving into more advanced topics. We start by going over the existing libraries for data processing and libraries with machine learning algorithms. After that, we cover topics such as classification and regression, dimensionality reduction and clustering, information retrieval and natural language processing, and deep learning and big data. Finally, we finish the course by talking about the ways to deploy the model and evaluate it in production settings.

Working in a hands-on learning environment, led by our Java for Data Science expert instructor, students will learn about and explore:

* An overview of modern Data Science and Machine Learning libraries available in Java
* Coverage of a broad set of topics, going from the basics of Machine Learning to Deep Learning and Big Data frameworks.
* Easy-to-follow illustrations and the running example of building a search engine.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get a solid understanding of the data processing toolbox available in Java
* Explore the Data Science ecosystem available in Java
* Find out how to approach different Machine Learning problems with Java
* Process unstructured information such as natural language text or images
* Create your own search engine
* Get state-of-the-art performance with XGBoost
* Learn how to build deep neural networks with DeepLearning4j
* Build applications that scale and process large amounts of data
* Deploy data science models to production and evaluate their performance

Audience & Pre-Requisites

This course is designed for beginners who wants to use Java to create a diverse range of Data Science applications and bring Data Science into production

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Data Science Using Java**

* Data Science Using Java
* Data science
* Data science process models
* Data science in Java

1. **Data Processing Toolbox**

* Data Processing Toolbox
* Standard Java library
* Extensions to the standard library
* Accessing data
* Search engine - preparing data

1. **Exploratory Data Analysis**

* Exploratory Data Analysis
* Exploratory data analysis in Java
* Interactive Exploratory Data Analysis in Java

1. **Supervised Learning - Classification and Regression**

* Supervised Learning - Classification and Regression
* Classification
* Case study - page prediction
* Regression
* Case study - hardware performance

1. **Unsupervised Learning - Clustering and Dimensionality Reduction**

* Unsupervised Learning - Clustering and Dimensionality Reduction
* Dimensionality reduction
* Cluster analysis

1. **Working with Text - Natural Language Processing and Information Retrieval**

* Working with Text - Natural Language Processing and Information Retrieval
* Natural Language Processing and information retrieval
* Machine learning for texts

1. **Extreme Gradient Boosting**

* Extreme Gradient Boosting
* Gradient Boosting Machines and XGBoost
* XGBoost in practice

1. **Deep Learning with DeepLearning4J**

* Deep Learning with DeepLearning4J
* Neural Networks and DeepLearning4J
* Deep learning for cats versus dogs

1. **caling Data Science**

* Scaling Data Science
* Apache Hadoop
* Apache Spark
* Link prediction
* Summary
* 10Deploying Data Science Models
* Deploying Data Science Models
* Microservices
* Online evaluation

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Python Social Media Analytics

Course Snapshot

* **Course:** Python Social Media Analytics
* **Duration:** 2 days
* **Skill-level**: Foundation-level Python Social Media skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Leverage the power of Python to collect, process, and mine deep insights from social media data
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Social Media platforms such as Facebook, Twitter, Forums, Pinterest, and YouTube have become part of everyday life in a big way. However, these complex and noisy data streams pose a potent challenge to everyone when it comes to harnessing them properly and benefiting from them. This book will introduce you to the concept of social media analytics, and how you can leverage its capabilities to empower your business. Right from acquiring data from various social networking sources such as Twitter, Facebook, YouTube, Pinterest, and social forums, you will see how to clean data and make it ready for analytical operations using various Python APIs. This book explains how to structure the clean data obtained and store in MongoDB using PyMongo. You will also perform web scraping and visualize data using Scrappy and Beautiful soup. Finally, you will be introduced to different techniques to perform analytics at scale for your social data on the cloud, using Python and Spark. By the end of this course, you will be able to utilize the power of Python to gain valuable insights from social media data and use them to enhance your business processes.

Working in a hands-on learning environment, led by our Python Social Media expert instructor, students will learn about and explore:

* Acquire data from various social media platforms such as Facebook, Twitter, YouTube, GitHub, and more
* Analyze and extract actionable insights from your social data using various Python tools
* A highly practical guide to conducting efficient social media analytics at scale

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the basics of social media mining
* Use PyMongo to clean, store, and access data in MongoDB
* Understand user reactions and emotion detection on Facebook
* Perform Twitter sentiment analysis and entity recognition using Python
* Analyze video and campaign performance on YouTube
* Mine popular trends on GitHub and predict the next big technology
* Extract conversational topics on public internet forums
* Analyze user interests on Pinterest
* Perform large-scale social media analytics on the cloud

Audience & Pre-Requisites

This course is designed for for beginners who know the Leverage the power of Python to collect, process, and mine deep insights from social media data

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Introduction to the Latest Social Media Landscape and Importance**

* Introduction to the Latest Social Media Landscape and Importance
* Introducing social graph
* Delving into social data
* Understanding the process
* Working environment
* Getting the data
* Analyzing the data
* Visualizing the data
* Getting started with the toolset

1. **Harnessing Social Data - Connecting, Capturing, and Cleaning**

* Harnessing Social Data - Connecting, Capturing, and Cleaning
* APIs in a nutshell
* Introduction to authentication techniques
* Parsing API outputs
* Basic cleaning techniques
* MongoDB to store and access social data
* MongoDB using Python

1. **Uncovering Brand Activity, Popularity, and Emotions on Facebook**

* Uncovering Brand Activity, Popularity, and Emotions on Facebook
* Facebook brand page
* Project planning
* Analysis
* Keywords
* Noun phrases
* Detecting trends in time series
* Uncovering emotions
* How can brands benefit from it?

1. **Analyzing Twitter Using Sentiment Analysis and Entity Recognition**

* Analyzing Twitter Using Sentiment Analysis and Entity Recognition
* Scope and process
* Getting the data
* Sentiment analysis
* Customized sentiment analysis
* Named entity recognition
* Combining NER and sentiment analysis

1. **Campaigns and Consumer Reaction Analytics on YouTube – Structured and Unstructured**

* Campaigns and Consumer Reaction Analytics on YouTube – Structured and Unstructured
* Scope and process
* Getting the data
* Data pull
* Data processing
* Data analysis

1. **The Next Great Technology – Trends Mining on GitHub**

* The Next Great Technology – Trends Mining on GitHub
* Scope and process
* Getting the data
* Data pull
* Data processing
* Data analysis

1. **Scraping and Extracting Conversational Topics on** **Internet Forums**

* Scraping and Extracting Conversational Topics on Internet Forums
* Scope and process
* Getting the data
* Data pull and pre-processing
* Data analysis

1. **Demystifying Pinterest through Network Analysis of Users Interests**

* Demystifying Pinterest through Network Analysis of Users Interests
* Scope and process
* Getting the data
* Data pull and pre-processing
* Data analysis

1. **Social Data Analytics at Scale – Spark and Amazon Web Services**

* Social Data Analytics at Scale – Spark and Amazon Web Services
* Different scaling methods and platforms
* Topic models at scale
* Spark on the Cloud â“ Amazon Elastic MapReduce

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Learning Social Media Analytics with R

Course Snapshot

* **Course:** Learning Social Media Analytics with R
* **Duration:** 2 days
* **Skill-level**: Foundation-level Social Media skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Tap into the realm of social media and unleash the power of analytics for data-driven insights using R
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The Internet has truly become humongous, especially with the rise of various forms of social media in the last decade, which give users a platform to express themselves and also communicate and collaborate with each other. This course will help the reader to understand the current social media landscape and to learn how analytics can be leveraged to derive insights from it. This data can be analyzed to gain valuable insights into the behavior and engagement of users, organizations, businesses, and brands. It will help readers frame business problems and solve them using social data. The course will also cover several practical real-world use cases on social media using R and its advanced packages to utilize data science methodologies such as sentiment analysis, topic modeling, text summarization, recommendation systems, social network analysis, classification, and clustering. This will enable readers to learn different hands-on approaches to obtain data from diverse social media sources such as Twitter and Facebook. It will also show readers how to establish detailed workflows to process, visualize, and analyze data to transform social data into actionable insights.

Working in a hands-on learning environment, led by our Social Media Analytics with R expert instructor, students will learn about and explore:

* A practical guide written to help leverage the power of the R eco-system to extract, process, analyze, visualize and model social media data
* Learn about data access, retrieval, cleaning, and curation methods for data originating from various social media platforms.
* Visualize and analyze data from social media platforms to understand and model complex relationships using various concepts and techniques such as Sentiment Analysis, Topic Modeling, Text Summarization, Recommendation Systems, Social Network Analysis, Classification, and Clustering.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn how to tap into data from diverse social media platforms using the R ecosystem
* Use social media data to formulate and solve real-world problems
* Analyze user social networks and communities using concepts from graph theory and network analysis
* Learn to detect opinion and sentiment, extract themes, topics, and trends from unstructured noisy text data from diverse social media channels
* Understand the art of representing actionable insights with effective visualizations
* Analyze data from major social media channels such as Twitter, Facebook, Flickr, Foursquare, Github, StackExchange, and so on
* Learn to leverage popular R packages such as ggplot2, topicmodels, caret, e1071, tm, wordcloud, twittR, Rfacebook, dplyr, reshape2, and many more

Audience & Pre-Requisites

This course is designed for for beginners who wants to tap into the realm of social media and unleash the power of analytics for data-driven insights using R

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Getting Started with R and Social Media Analytics**

* Getting Started with R and Social Media Analytics
* Understanding social media
* Social media analytics
* Getting started with R
* Data types
* Data analytics
* Machine learning
* Text analytics

1. **Twitter – What's Happening with 140 Characters**

* Twitter – What's Happening with 140 Characters
* Understanding Twitter
* Revisiting analytics workflow
* Trend analysis
* Sentiment analysis
* Follower graph analysis

1. **Analyzing Social Networks and Brand Engagements with Facebook**

* Analyzing Social Networks and Brand Engagements with Facebook
* Accessing Facebook data
* Analyzing your personal social network
* Analyzing an English football social network
* Analyzing English Football Club's brand page engagements

1. **Foursquare – Are You Checked in Yet?**

* Foursquare – Are You Checked in Yet?
* Foursquare – the app and data
* Category trend analysis
* Recommendation engine – let's open a restaurant
* The sentimental rankings
* Venue graph – where do people go next?
* Challenges for Foursquare data analysis

1. **Analyzing Software Collaboration Trends I – Social Coding with GitHub**

* Analyzing Software Collaboration Trends I – Social Coding with GitHub
* Environment setup
* Understanding GitHub
* Accessing GitHub data
* Analyzing repository activity
* Analyzing repository trends
* Analyzing language trends

1. **Analyzing Software Collaboration Trends II - Answering Your Questions** **with StackExchange**

* Analyzing Software Collaboration Trends II - Answering Your Questions with StackExchange
* Understanding StackExchange
* Data Science and StackExchange
* Demographics and data science
* Challenges

1. **Believe What You See – Flickr Data Analysis**

* Believe What You See – Flickr Data Analysis
* A Flickr-ing world
* Accessing Flickr's data
* Understanding Flickr data
* Understanding interestingness – similarities
* Are your photos interesting?
* Challenges

1. **News – The Collective Social Media!**

* News – The Collective Social Media!
* News data – news is everywhere
* Sentiment trend analysis
* Topic modeling
* Summarizing news articles
* Challenges to news data analysis

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work

# Python for Finance

Course Snapshot

* **Course:** Python for Finance
* **Duration:** 4 days
* **Skill-level**: Foundation-level Python for Finance skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Learn and implement various Quantitative Finance concepts using the popular Python libraries
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

This course uses Python as its computational tool. Since Python is free, any school or organization can download and use it. This course is organized according to various finance subjects. In other words, the first edition focuses more on Python, while the this edition is truly trying to apply Python to finance. The course starts by explaining topics exclusively related to Python. Then we deal with critical parts of Python, explaining concepts such as time value of money stock and bond evaluations, capital asset pricing model, multi-factor models, time series analysis, portfolio theory, options and futures. This course will help us to learn or review the basics of quantitative finance and apply Python to solve various problems, such as estimating IBM’s market risk, running a Fama-French 3-factor, 5-factor, or Fama-French-Carhart 4 factor model, estimating the VaR of a 5-stock portfolio, estimating the optimal portfolio, and constructing the efficient frontier for a 20-stock portfolio with real-world stock, and with Monte Carlo Simulation. Later, we will also learn how to replicate the famous Black-Scholes-Merton option model and how to price exotic options such as the average price call option.

Working in a hands-on learning environment, led by our Python for Finance expert instructor, students will learn about and explore:

* Understand the fundamentals of Python data structures and work with time-series data
* Implement key concepts in quantitative finance using popular Python libraries such as NumPy, SciPy, and matplotlib
* A step-by-step tutorial packed with many Python programs that will help you learn how to apply Python to finance

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Become acquainted with Python in the first two chapters
* Run CAPM, Fama-French 3-factor, and Fama-French-Carhart 4-factor models
* Learn how to price a call, put, and several exotic options
* Understand Monte Carlo simulation, how to write a Python program to replicate the Black-Scholes-Merton options model, and how to price a few exotic options
* Understand the concept of volatility and how to test the hypothesis that volatility changes over the years
* Understand the ARCH and GARCH processes and how to write related Python programs

Audience & Pre-Requisites

This course is designed for for beginners who wants to learn and implement various Quantitative Finance concepts using the popular Python libraries.

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Python Basics**

* Python Basics
* Python installation
* Variable assignment, empty space, and writing our own programs
* Writing a Python function
* Python loops
* Data input
* Data manipulation
* Data output

1. **Introduction to Python Modules**

* Introduction to Python Modules
* What is a Python module?
* Introduction to NumPy
* Introduction to SciPy
* Introduction to matplotlib
* Introduction to statsmodels
* Introduction to pandas
* Python modules related to finance
* Introduction to the pandas\_reader module
* Two financial calculators
* How to install a Python module
* Module dependency

1. **Time Value of Money**

* Time Value of Money
* Introduction to time value of money
* Writing a financial calculator in Python
* Definition of NPV and NPV rule
* Definition of IRR and IRR rule
* Definition of payback period and payback period rule
* Writing your own financial calculator in Python
* Two general formulae for many functions

1. **Sources of Data**

* Sources of Data
* Diving into deeper concepts

1. **Bond and Stock** V**aluation**

* Bond and Stock Valuation
* Introduction to interest rates
* Term structure of interest rates
* Bond evaluation
* Stock valuation
* A new data type – dictionary

1. **Capital Asset Pricing Model**

* Capital Asset Pricing Model
* Introduction to CAPM
* Moving beta
* Adjusted beta
* Extracting output data
* Simple string manipulation
* Python via Canopy

1. **Multifactor Models and Performance Measures**

* Multifactor Models and Performance Measures
* Introduction to the Fama-French three-factor model
* Fama-French three-factor model
* Fama-French-Carhart four-factor model and Fama-French five-factor model
* Implementation of Dimson (1979) adjustment for beta
* Performance measures
* How to merge different datasets

1. **Time-Series Analysis**

* Time-Series Analysis
* Introduction to time-series analysis
* Merging datasets based on a date variable
* Understanding the interpolation technique
* Tests of normality
* 52-week high and low trading strategy
* Estimating Roll's spread
* Estimating Amihud's illiquidity
* Estimating Pastor and Stambaugh (2003) liquidity measure
* Fama-MacBeth regression
* Durbin-Watson
* Python for high-frequency data
* Spread estimated based on high-frequency data
* Introduction to CRSP

1. **Portfolio Theory**

* Portfolio Theory
* Introduction to portfolio theory
* A 2-stock portfolio
* Optimization – minimization
* Forming an n-stock portfolio
* Constructing an optimal portfolio
* Constructing an efficient frontier with n stocks

1. **Options and Futures**

* Options and Futures
* Introducing futures
* Payoff and profit/loss functions for call and put options
* European versus American options
* Black-Scholes-Merton option model on non-dividend paying stocks
* Generating our own module p4f
* European options with known dividends
* Various trading strategies
* Put-call parity and its graphic presentation
* Binomial tree and its graphic presentation
* Hedging strategies
* Implied volatility
* Binary-search
* Retrieving option data from Yahoo! Finance
* Volatility smile and skewness

1. **Value at Risk**

* Value at Risk
* Introduction to VaR
* Normality tests
* Skewness and kurtosis
* Modified VaR
* VaR based on sorted historical returns
* Simulation and VaR
* VaR for portfolios
* Backtesting and stress testing
* Expected shortfall

1. **Monte Carlo Simulation**

* Monte Carlo Simulation
* Importance of Monte Carlo Simulation
* Generating random numbers from a standard normal distribution
* Generating random numbers with a seed
* Generating random numbers from a uniform distribution
* Using simulation to estimate the pi value
* Generating random numbers from a Poisson distribution
* Selecting m stocks randomly from n given stocks
* With/without replacements
* Distribution of annual returns
* Simulation of stock price movements
* Graphical presentation of stock prices at options' maturity dates
* Replicating a Black-Scholes-Merton call using simulation
* Liking two methods for VaR using simulation
* Capital budgeting with Monte Carlo Simulation
* Python SimPy module
* Comparison between two social policies – basic income and basic job
* Finding an efficient frontier based on two stocks by using simulation
* Constructing an efficient frontier with n stocks
* Long-term return forecasting
* Efficiency, Quasi-Monte Carlo, and Sobol sequences

1. **Credit Risk Analysis**

* Credit Risk Analysis
* Introduction to credit risk analysis
* Credit rating
* Credit spread
* YIELD of AAA-rated bond, Altman Z-score
* Using the KMV model to estimate the market value of total assets and its volatility
* Term structure of interest rate
* Distance to default
* Credit default swap

1. **Exotic Options**

* Exotic Options
* European, American, and Bermuda options
* Chooser options
* Shout options
* Binary options
* Rainbow options
* Pricing average options
* Pricing barrier options
* Barrier in-and-out parity
* Graph of up-and-out and up-and-in parity
* Pricing lookback options with floating strikes

1. **Volatility, Implied Volatility, ARCH, and GARCH**

* Volatility, Implied Volatility, ARCH, and GARCH
* Conventional volatility measure – standard deviation
* Tests of normality
* Estimating fat tails
* Lower partial standard deviation and Sortino ratio
* Test of equivalency of volatility over two periods
* Test of heteroskedasticity, Breusch, and Pagan
* Volatility smile and skewness
* Graphical presentation of volatility clustering
* The ARCH model
* Simulating an ARCH (1) process
* The GARCH model
* Simulating a GARCH process
* Simulating a GARCH (p,q) process using modified garchSim()
* GJR\_GARCH by Glosten, Jagannanthan, and Runkle

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Real-time Data Processing and Analytics

Course Snapshot

* **Course:** Real-time Data Processing and Analytics
* **Duration:** 3 days
* **Skill-level**: Foundation-level Practical Real-time Data Processing and Analytics skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Get practical guide to help you tackle different real-time data processing and analytics problems using the best tools for each scenario.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

With the rise of Big Data, there is an increasing need to process large amounts of data continuously, with a shorter turnaround time. Real-time data processing involves continuous input, processing and output of data, with the condition that the time required for processing is as short as possible. This book covers the majority of the existing and evolving open source technology stack for real-time processing and analytics. You will get to know about all the real-time solution aspects, from the source to the presentation to persistence. Through this practical book, you’ll be equipped with a clear understanding of how to solve challenges on your own. We’ll cover topics such as how to set up components, basic executions, integrations, advanced use cases, alerts, and monitoring. You’ll be exposed to the popular tools used in real-time processing today such as Apache Spark, Apache Flink, and Storm. Finally, you will put your knowledge to practical use by implementing all of the techniques in the form of a practical, real-world use case. By the end of this book, you will have a solid understanding of all the aspects of real-time data processing and analytics, and will know how to deploy the solutions in production environments in the best possible manner.

Working in a hands-on learning environment, led by our Real-time Data Processing and Analytics expert instructor, students will learn about and explore:

* Learn about the various challenges in real-time data processing and use the right tools to overcome them
* This book covers popular tools and frameworks such as Spark, Flink, and Apache Storm to solve all your distributed processing problems
* A practical guide filled with examples, tips, and tricks to help you perform efficient Big Data processing in real-time

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get an introduction to the established real-time stack
* Understand the key integration of all the components
* Get a thorough understanding of the basic building blocks for real-time solution designing
* Garnish the search and visualization aspects for your real-time solution
* Get conceptually and practically acquainted with real-time analytics
* Be well equipped to apply the knowledge and create your own solutions

Audience & Pre-Requisites

This course is designed for for beginners who wants to get practical guide to help you tackle different real-time data processing and analytics problems using the best tools for each scenario

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Introducing Real-Time Analytics**

* Introducing Real-Time Analytics
* What is big data?
* Big data infrastructure
* Real–time analytics – the myth and the reality
* Near real–time solution – an architecture that works
* Lambda architecture – analytics possibilities
* IOT – thoughts and possibilities
* Cloud – considerations for NRT and IOT

1. **Real Time Applications –** **The Basic Ingredients**

* Real Time Applications – The Basic Ingredients
* The NRT system and its building blocks
* NRT – high-level system view
* NRT – technology view

1. **Understanding and Tailing Data Streams**

* Understanding and Tailing Data Streams
* Understanding data streams
* Setting up infrastructure for data ingestion
* Taping data from source to the processor - expectations and caveats
* Comparing and choosing what works best for your use case
* Do it yourself

1. **Setting up the** **Infrastructure for Storm**

* Setting up the Infrastructure for Storm
* Overview of Storm
* Storm architecture and its components
* Setting up and configuring Storm
* Real-time processing job on Storm

1. **Configuring Apache Spark and Flink**

* Configuring Apache Spark and Flink
* Setting up and a quick execution of Spark
* Setting up and a quick execution of Flink
* Setting up and a quick execution of Apache Beam
* Balancing in Apache Beam

1. **Integrating Storm with a Data Source**

* Integrating Storm with a Data Source
* RabbitMQ – messaging that works
* RabbitMQ exchanges
* RabbitMQ – integration with Storm
* PubNub data stream publisher
* String together Storm-RMQ-PubNub sensor data topology

1. **From Storm to Sink**

* From Storm to Sink
* Setting up and configuring Cassandra
* Storm and Cassandra topology
* Storm and IMDB integration for dimensional data
* Integrating the presentation layer with Storm
* Do It Yourself

1. **Storm Trident**

* Storm Trident
* State retention and the need for Trident
* Basic Storm Trident topology
* Trident internals
* Trident operations
* DRPC
* Do It Yourself

1. **Working with Spark**

* Working with Spark
* Spark overview
* Distinct advantages of Spark
* Spark – use cases
* Spark architecture - working inside the engine
* Spark pragmatic concepts
* Spark 2.x – advent of data frames and datasets

1. **Working with Spark Operations**

* Working with Spark Operations
* Spark – packaging and API
* RDD pragmatic exploration
* Shared variables – broadcast variables and accumulators

1. **Spark Streaming**

* Spark Streaming
* Spark Streaming concepts
* Spark Streaming - introduction and architecture
* Packaging structure of Spark Streaming
* Connecting Kafka to Spark Streaming

1. **Working with Apache Flink**

* Working with Apache Flink
* Flink architecture and execution engine
* Flink basic components and processes
* Integration of source stream to Flink
* Flink processing and computation
* Flink persistence
* FlinkCEP
* Pattern API
* Gelly
* DIY

1. **Case Study**

* Case Study
* Introduction
* Data modeling
* Tools and frameworks
* Setting up the infrastructure
* Implementing the case study
* Running the case study

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Apache Spark 2.x Machine Learning Cookbook

Course Snapshot

* **Course:** Apache Spark 2.x Machine Learning Cookbook
* **Duration:** 3 days
* **Skill-level**: Foundation-level Apache Spark 2.x Machine Learning Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Simplify machine learning model implementations with Spark.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning aims to extract knowledge from data, relying on fundamental concepts in computer science, statistics, probability, and optimization. Learning about algorithms enables a wide range of applications, from everyday tasks such as product recommendations and spam filtering to cutting edge applications such as self-driving cars and personalized medicine. You will gain hands-on experience of applying these principles using Apache Spark, a resilient cluster computing system well suited for large-scale machine learning tasks. This book begins with a quick overview of setting up the necessary IDEs to facilitate the execution of code examples that will be covered in various chapters. It also highlights some key issues developers face while working with machine learning algorithms on the Spark platform. We progress by uncovering the various Spark APIs and the implementation of ML algorithms with developing classification systems, recommendation engines, text analytics, clustering, and learning systems. Toward the final chapters, we’ll focus on building high-end applications and explain various unsupervised methodologies and challenges to tackle when implementing with big data ML systems.

Working in a hands-on learning environment, led by our Data Science with Python and Jupyter expert instructor, students will learn about and explore:

* Solve the day-to-day problems of data science with Spark
* This unique cookbook consists of exciting and intuitive numerical recipes
* Optimize your work by acquiring, cleaning, analyzing, predicting, and visualizing your data

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get to know how Scala and Spark go hand-in-hand for developers when developing ML systems with Spark
* Build a recommendation engine that scales with Spark
* Find out how to build unsupervised clustering systems to classify data in Spark
* Build machine learning systems with the Decision Tree and Ensemble models in Spark
* Deal with the curse of high-dimensionality in big data using Spark
* Implement Text analytics for Search Engines in Spark
* Streaming Machine Learning System implementation using Spark

Audience & Pre-Requisites

This course is designed for for beginners who wants to simplify machine learning model implementations with Spark.

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Practical Machine Learning with Spark Using Scala**

* Practical Machine Learning with Spark Using Scala
* Introduction
* Downloading and installing the JDK
* Downloading and installing IntelliJ
* Downloading and installing Spark
* Configuring IntelliJ to work with Spark and run Spark ML sample codes
* Running a sample ML code from Spark
* Identifying data sources for practical machine learning
* Running your first program using Apache Spark 2.0 with the IntelliJ IDE
* How to add graphics to your Spark program

1. **Just Enough Linear Algebra for Machine Learning with Spark**

* Just Enough Linear Algebra for Machine Learning with Spark
* Introduction
* Package imports and initial setup for vectors and matrices
* Creating DenseVector and setup with Spark 2.0
* Creating SparseVector and setup with Spark
* Creating dense matrix and setup with Spark 2.0
* Using sparse local matrices with Spark 2.0
* Performing vector arithmetic using Spark 2.0
* Performing matrix arithmetic using Spark 2.0
* Exploring RowMatrix in Spark 2.0
* Exploring Distributed IndexedRowMatrix in Spark 2.0
* Exploring distributed CoordinateMatrix in Spark 2.0
* Exploring distributed BlockMatrix in Spark 2.0

1. **Spark's Three Data Musketeers for Machine Learning - Perfect Together**

* Spark's Three Data Musketeers for Machine Learning - Perfect Together
* Introduction
* Creating RDDs with Spark 2.0 using internal data sources
* Creating RDDs with Spark 2.0 using external data sources
* Transforming RDDs with Spark 2.0 using the filter() API
* Transforming RDDs with the super useful flatMap() API
* Transforming RDDs with set operation APIs
* RDD transformation/aggregation with groupBy() and reduceByKey()
* Transforming RDDs with the zip() API
* Join transformation with paired key-value RDDs
* Reduce and grouping transformation with paired key-value RDDs
* Creating DataFrames from Scala data structures
* Operating on DataFrames programmatically without SQL
* Loading DataFrames and setup from an external source
* Using DataFrames with standard SQL language - SparkSQL
* Working with the Dataset API using a Scala Sequence
* Creating and using Datasets from RDDs and back again
* Working with JSON using the Dataset API and SQL together
* Functional programming with the Dataset API using domain objects

1. **Common Recipes for Implementing a Robust Machine Learning System**

* Common Recipes for Implementing a Robust Machine Learning System
* Introduction
* Spark's basic statistical API to help you build your own algorithms
* ML pipelines for real-life machine learning applications
* Normalizing data with Spark
* Splitting data for training and testing
* Common operations with the new Dataset API
* Creating and using RDD versus DataFrame versus Dataset from a text file in Spark 2.0
* LabeledPoint data structure for Spark ML
* Getting access to Spark cluster in Spark 2.0
* Getting access to Spark cluster pre-Spark 2.0
* Getting access to SparkContext vis-a-vis SparkSession object in Spark 2.0
* New model export and PMML markup in Spark 2.0
* Regression model evaluation using Spark 2.0
* Binary classification model evaluation using Spark 2.0
* Multiclass classification model evaluation using Spark 2.0
* Multilabel classification model evaluation using Spark 2.0
* Using the Scala Breeze library to do graphics in Spark 2.0

1. **Practical Machine Learning with Regression and Classification in Spark 2.0 - Part I**

* Practical Machine Learning with Regression and Classification in Spark 2.0 - Part I
* Introduction
* Fitting a linear regression line to data the old fashioned way
* Generalized linear regression in Spark 2.0
* Linear regression API with Lasso and L-BFGS in Spark 2.0
* Linear regression API with Lasso and 'auto' optimization selection in Spark 2.0
* Linear regression API with ridge regression and 'auto' optimization selection in Spark 2.0
* Isotonic regression in Apache Spark 2.0
* Multilayer perceptron classifier in Apache Spark 2.0
* One-vs-Rest classifier (One-vs-All) in Apache Spark 2.0
* Survival regression – parametric AFT model in Apache Spark 2.0

1. **Practical Machine Learning with Regression and Classification in Spark 2.0 - Part II**

* Practical Machine Learning with Regression and Classification in Spark 2.0 - Part II
* Introduction
* Linear regression with SGD optimization in Spark 2.0
* Logistic regression with SGD optimization in Spark 2.0
* Ridge regression with SGD optimization in Spark 2.0
* Lasso regression with SGD optimization in Spark 2.0
* Logistic regression with L-BFGS optimization in Spark 2.0
* Support Vector Machine (SVM) with Spark 2.0
* Naive Bayes machine learning with Spark 2.0 MLlib
* Exploring ML pipelines and DataFrames using logistic regression in Spark 2.0

1. **Recommendation Engine that Scales with Spark**

* Recommendation Engine that Scales with Spark
* Introduction
* Setting up the required data for a scalable recommendation engine in Spark 2.0
* Exploring the movies data details for the recommendation system in Spark 2.0
* Exploring the ratings data details for the recommendation system in Spark 2.0
* Building a scalable recommendation engine using collaborative filtering in Spark 2.0

1. **Unsupervised Clustering with Apache Spark 2.0**

* Unsupervised Clustering with Apache Spark 2.0
* Introduction
* Building a KMeans classifying system in Spark 2.0
* Bisecting KMeans, the new kid on the block in Spark 2.0
* Using Gaussian Mixture and Expectation Maximization (EM) in Spark to classify data
* Classifying the vertices of a graph using Power Iteration Clustering (PIC) in Spark 2.0
* Latent Dirichlet Allocation (LDA) to classify documents and text into topics
* Streaming KMeans to classify data in near real-time

1. **Optimization - Going Down the Hill with Gradient Descent**

* Optimization - Going Down the Hill with Gradient Descent
* Introduction
* Optimizing a quadratic cost function and finding the minima using just math to gain insight
* Coding a quadratic cost function optimization using Gradient Descent (GD) from scratch
* Coding Gradient Descent optimization to solve Linear Regression from scratch
* Normal equations as an alternative for solving Linear Regression in Spark 2.0

1. **Building Machine Learning Systems with Decision Tree and Ensemble Models**

* Building Machine Learning Systems with Decision Tree and Ensemble Models
* Introduction
* Getting and preparing real-world medical data for exploring Decision Trees and Ensemble models in Spark 2.0
* Building a classification system with Decision Trees in Spark 2.0
* Solving Regression problems with Decision Trees in Spark 2.0
* Building a classification system with Random Forest Trees in Spark 2.0
* Solving regression problems with Random Forest Trees in Spark 2.0
* Building a classification system with Gradient Boosted Trees (GBT) in Spark 2.0
* Solving regression problems with Gradient Boosted Trees (GBT) in Spark 2.0

1. **Curse of High-Dimensionality in Big Data**

* Curse of High-Dimensionality in Big Data
* Introduction
* Two methods of ingesting and preparing a CSV file for processing in Spark
* Singular Value Decomposition (SVD) to reduce high-dimensionality in Spark
* Principal Component Analysis (PCA) to pick the most effective latent factor for machine learning in Spark

1. **Implementing Text Analytics with Spark 2.0 ML Library**

* Implementing Text Analytics with Spark 2.0 ML Library
* Introduction
* Doing term frequency with Spark - everything that counts
* Displaying similar words with Spark using Word2Vec
* Downloading a complete dump of Wikipedia for a real-life Spark ML project
* Using Latent Semantic Analysis for text analytics with Spark 2.0
* Topic modeling with Latent Dirichlet allocation in Spark 2.0

1. **Spark Streaming and Machine Learning Library**

* Spark Streaming and Machine Learning Library
* Introduction
* Structured streaming for near real-time machine learning
* Streaming DataFrames for real-time machine learning
* Streaming Datasets for real-time machine learning
* Streaming data and debugging with queueStream
* Downloading and understanding the famous Iris data for unsupervised classification
* Streaming KMeans for a real-time on-line classifier
* Downloading wine quality data for streaming regression
* Streaming linear regression for a real-time regression
* Downloading Pima Diabetes data for supervised classification
* Streaming logistic regression for an on-line classifier

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Python Machine Learning Cookbook

Course Snapshot

* **Course:** Python Machine Learning Cookbook
* **Duration:** 4 days
* **Skill-level**: Foundation-level Python Machine Learning Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Discover powerful ways to effectively solve real-world machine learning problems using key libraries including scikit-learn, TensorFlow, and PyTorch.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

This eagerly anticipated edition of the popular Python Machine Learning Cookbook will enable you to adopt a fresh approach to dealing with real-world machine learning and deep learning tasks. With the help of over 100 recipes, you will learn to build powerful machine learning applications using modern libraries from the Python ecosystem. The course will also guide you on how to implement various machine learning algorithms for classification, clustering, and recommendation engines, using a recipe-based approach. With emphasis on practical solutions, dedicated sections in the course will help you to apply supervised and unsupervised learning techniques to real-world problems. Toward the concluding lessons, you will get to grips with recipes that teach you advanced techniques including reinforcement learning, deep neural networks, and automated machine learning. By the end of this course, you will be equipped with the skills you need to apply machine learning techniques and leverage the full capabilities of the Python ecosystem through real-world examples.

Working in a hands-on learning environment, led by our Python Machine Learning Cookbook expert instructor, students will learn about and explore:

* Learn and implement machine learning algorithms in a variety of real-life scenarios
* Cover a range of tasks catering to supervised, unsupervised and reinforcement learning techniques
* Find easy-to-follow code solutions for tackling common and not-so-common challenges

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use predictive modeling and apply it to real-world problems
* Explore data visualization techniques to interact with your data
* Learn how to build a recommendation engine
* Understand how to interact with text data and build models to analyze it
* Work with speech data and recognize spoken words using Hidden Markov Models
* Get well versed with reinforcement learning, automated ML, and transfer learning
* Work with image data and build systems for image recognition and biometric face recognition
* Use deep neural networks to build an optical character recognition system

Audience & Pre-Requisites

This course is designed for developers wants to Discover powerful ways to effectively solve real-world machine learning problems using key libraries including scikit-learn, TensorFlow, and PyTorch

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **The Realm of Supervised Learning**

* The Realm of Supervised Learning
* Technical requirements
* Introduction
* Array creation in Python
* Data preprocessing using mean removal
* Data scaling
* Normalization
* Binarization
* One-hot encoding
* Label encoding
* Building a linear regressor
* Computing regression accuracy
* Achieving model persistence
* Building a ridge regressor
* Building a polynomial regressor
* Estimating housing prices
* Computing the relative importance of features
* Estimating bicycle demand distribution

1. **Constructing a Classifier**

* Constructing a Classifier
* Technical requirements
* Introduction
* Building a simple classifier
* Building a logistic regression classifier
* Building a Naive Bayes classifier
* Splitting a dataset for training and testing
* Evaluating accuracy using cross-validation metrics
* Visualizing a confusion matrix
* Extracting a performance report
* Evaluating cars based on their characteristics
* Extracting validation curves
* Extracting learning curves
* Estimating the income bracket
* Predicting the quality of wine
* Newsgroup trending topics classification

1. **Predictive Modeling**

* Predictive Modeling
* Technical requirements
* Introduction
* Building a linear classifier using SVMs
* Building a nonlinear classifier using SVMs
* Tackling class imbalance
* Extracting confidence measurements
* Finding optimal hyperparameters
* Building an event predictor
* Estimating traffic
* Simplifying machine learning workflow using TensorFlow
* Implementing a stacking method

1. **Clustering with Unsupervised Learning**

* Clustering with Unsupervised Learning
* Technical requirements
* Introduction
* Clustering data using the k-means algorithm
* Compressing an image using vector quantization
* Grouping data using agglomerative clustering
* Evaluating the performance of clustering algorithms
* Estimating the number of clusters using the DBSCAN algorithm
* Finding patterns in stock market data
* Building a customer segmentation model
* Using autoencoders to reconstruct handwritten digit images

1. **Visualizing Data**

* Visualizing Data
* Technical requirements
* An introduction to data visualization
* Plotting three-dimensional scatter plots
* Plotting bubble plots
* Animating bubble plots
* Drawing pie charts
* Plotting date-formatted time series data
* Plotting histograms
* Visualizing heat maps
* Animating dynamic signals
* Working with the Seaborn library

1. **Building Recommendation Engines**

* Building Recommendation Engines
* Technical requirements
* Introducing the recommendation engine
* Building function compositions for data processing
* Building machine learning pipelines
* Finding the nearest neighbors
* Constructing a k-nearest neighbors classifier
* Constructing a k-nearest neighbors regressor
* Computing the Euclidean distance score
* Computing the Pearson correlation score
* Finding similar users in the dataset
* Generating movie recommendations
* Implementing ranking algorithms
* Building a filtering model using TensorFlow

1. **Analyzing Text Data**

* Analyzing Text Data
* Technical requirements
* Introduction
* Preprocessing data using tokenization
* Stemming text data
* Converting text to its base form using lemmatization
* Dividing text using chunking
* Building a bag-of-words model
* Building a text classifier
* Identifying the gender of a name
* Analyzing the sentiment of a sentence
* Identifying patterns in text using topic modeling
* Parts of speech tagging with spaCy
* Word2Vec using gensim
* Shallow learning for spam detection

1. **Speech Recognition**

* Speech Recognition
* Technical requirements
* Introducing speech recognition
* Reading and plotting audio data
* Transforming audio signals into the frequency domain
* Generating audio signals with custom parameters
* Synthesizing music
* Extracting frequency domain features
* Building HMMs
* Building a speech recognizer
* Building a TTS system

1. **Dissecting Time Series and Sequential Data**

* Dissecting Time Series and Sequential Data
* Technical requirements
* Introducing time series
* Transforming data into a time series format
* Slicing time series data
* Operating on time series data
* Extracting statistics from time series data
* Building HMMs for sequential data
* Building CRFs for sequential text data
* Analyzing stock market data
* Using RNNs to predict time series data

1. **Analyzing Image Content**

* Analyzing Image Content
* Technical requirements
* Introducing computer vision
* Operating on images using OpenCV-Python
* Detecting edges
* Histogram equalization
* Detecting corners
* Detecting SIFT feature points
* Building a Star feature detector
* Creating features using Visual Codebook and vector quantization
* Training an image classifier using Extremely Random Forests
* Building an object recognizer
* Using Light GBM for image classification

1. **Biometric Face Recognition**

* Biometric Face Recognition
* Technical requirements
* Introduction
* Capturing and processing video from a webcam
* Building a face detector using Haar cascades
* Building eye and nose detectors
* Performing principal component analysis
* Performing kernel principal component analysis
* Performing blind source separation
* Building a face recognizer using a local binary patterns histogram
* Recognizing faces using the HOG-based model
* Facial landmark recognition
* User authentication by face recognition

1. **Reinforcement Learning Techniques**

* Reinforcement Learning Techniques
* Technical requirements
* Introduction
* Weather forecasting with MDP
* Optimizing a financial portfolio using DP
* Finding the shortest path
* Deciding the discount factor using Q-learning
* Implementing the deep Q-learning algorithm
* Developing an AI-based dynamic modeling system
* Deep reinforcement learning with double Q-learning
* Deep Q-network algorithm with dueling Q-learning

1. **Deep Neural Networks**

* Deep Neural Networks
* Technical requirements
* Introduction
* Building a perceptron
* Building a single layer neural network
* Building a deep neural network
* Creating a vector quantizer
* Building a recurrent neural network for sequential data analysis
* Visualizing the characters in an OCR database
* Building an optical character recognizer using neural networks
* Implementing optimization algorithms in ANN

1. **Unsupervised Representation Learning**

* Unsupervised Representation Learning
* Technical requirements
* Introduction
* Using denoising autoencoders to detect fraudulent transactions
* Generating word embeddings using CBOW and skipgram representations
* Visualizing the MNIST dataset using PCA and t-SNE
* Using word embedding for Twitter sentiment analysis
* Implementing LDA with scikit-learn
* Using LDA to classify text documents
* Preparing data for LDA

1. **Automated Machine Learning and Transfer Learning**

* Automated Machine Learning and Transfer Learning
* Technical requirements
* Introduction
* Working with Auto-WEKA
* Using AutoML to generate machine learning pipelines with TPOT
* Working with Auto-Keras
* Working with auto-sklearn
* Using MLBox for selection and leak detection
* Convolutional neural networks with transfer learning
* Transfer learning with pretrained image classifiers using ResNet-50
* Transfer learning using feature extraction with the VGG16 model
* Transfer learning with pretrained GloVe embedding

1. **Unlocking Production Issues**

* Unlocking Production Issues
* Technical requirements
* Introduction
* Handling unstructured data
* Deploying machine learning models
* Keeping track of changes into production
* Tracking accuracy to optimize model scaling

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# TensorFlow Machine Learning Cookbook

Course Snapshot

* **Course:** TensorFlow Machine Learning Cookbook
* **Duration:** 3 days
* **Skill-level**: Foundation-level TensorFlow Machine Learning Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Skip the theory and get the most out of Tensorflow to build production-ready machine learning models.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

TensorFlow is an open source software library for Machine Intelligence. The independent recipes in this course will teach you how to use TensorFlow for complex data computations and allow you to dig deeper and gain more insights into your data than ever before. With the help of this course, you will work with recipes for training models, model evaluation, sentiment analysis, regression analysis, clustering analysis, artificial neural networks, and more. You will explore RNNs, CNNs, GANs, reinforcement learning, and capsule networks, each using Google's machine learning library, TensorFlow. Through real-world examples, you will get hands-on experience with linear regression techniques with TensorFlow. Once you are familiar and comfortable with the TensorFlow ecosystem, you will be shown how to take it to production. By the end of the course, you will be proficient in the field of machine intelligence using TensorFlow. You will also have good insight into deep learning and be capable of implementing machine learning algorithms in real-world scenarios.

Working in a hands-on learning environment, led by our TensorFlow Machine Learning Cookbook expert instructor, students will learn about and explore:

* Exploit the features of Tensorflow to build and deploy machine learning models
* Train neural networks to tackle real-world problems in Computer Vision and NLP
* Handy techniques to write production-ready code for your Tensorflow models

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Become familiar with the basic features of the TensorFlow library
* Get to know Linear Regression techniques with TensorFlow
* Learn SVMs with hands-on recipes
* Implement neural networks to improve predictive modeling
* Apply NLP and sentiment analysis to your data
* Master CNN and RNN through practical recipes
* Implement the gradient boosted random forest to predict housing prices
* Take TensorFlow into production

Audience & Pre-Requisites

This course is designed for developers wants to Skip the theory and get the most out of Tensorflow to build production-ready machine learning models.

**Pre-Requisites:** Students should have familiar with

* Basics of ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Getting Started with TensorFlow**

* Getting Started with TensorFlow
* Introduction
* How TensorFlow works
* Declaring variables and tensors
* Using placeholders and variables
* Working with matrices
* Declaring operations
* Implementing activation functions
* Working with data sources
* Additional resources

1. **The TensorFlow Way**

* The TensorFlow Way
* Introduction
* Operations in a computational graph
* Layering nested operations
* Working with multiple layers
* Implementing loss functions
* Implementing backpropagation
* Working with batch and stochastic training
* Combining everything together
* Evaluating models

1. **Linear Regression**

* Linear Regression
* Introduction
* Using the matrix inverse method
* Implementing a decomposition method
* Learning the TensorFlow way of linear regression
* Understanding loss functions in linear regression
* Implementing deming regression
* Implementing lasso and ridge regression
* Implementing elastic net regression
* Implementing logistic regression

1. **Support Vector Machines**

* Support Vector Machines
* Introduction
* Working with a linear SVM
* Reduction to linear regression
* Working with kernels in TensorFlow
* Implementing a non-linear SVM
* Implementing a multi-class SVM

1. **Nearest-Neighbor Methods**

* Nearest-Neighbor Methods
* Introduction
* Working with nearest-neighbors
* Working with text based distances
* Computing with mixed distance functions
* Using an address matching example
* Using nearest-neighbors for image recognition

1. **Neural Networks**

* Neural Networks
* Introduction
* Implementing operational gates
* Working with gates and activation functions
* Implementing a one-layer neural network
* Implementing different layers
* Using a multilayer neural network
* Improving the predictions of linear models
* Learning to play Tic Tac Toe

1. **Natural Language Processing**

* Natural Language Processing
* Introduction
* Working with bag-of-words embeddings
* Implementing TF-IDF
* Working with Skip-Gram embeddings
* Working with CBOW embeddings
* Making predictions with word2vec
* Using doc2vec for sentiment analysis

1. **Convolutional Neural Networks**

* Convolutional Neural Networks
* Introduction
* Implementing a simple CNN
* Implementing an advanced CNN
* Retraining existing CNN models
* Applying stylenet and the neural-style project
* Implementing DeepDream

1. **Recurrent Neural Networks**

* Recurrent Neural Networks
* Introduction
* Implementing RNN for spam prediction
* Implementing an LSTM model
* Stacking multiple LSTM layers
* Creating sequence-to-sequence models
* Training a Siamese similarity measure

1. **Taking TensorFlow to Production**

* Taking TensorFlow to Production
* Introduction
* Implementing unit tests
* Using multiple executors
* Parallelizing TensorFlow
* Taking TensorFlow to production
* An example of productionalizing TensorFlow
* Using TensorFlow Serving

1. **More with TensorFlow**

* More with TensorFlow
* Introduction
* Visualizing graphs in TensorBoard
* Working with a genetic algorithm
* Clustering using k-means
* Solving a system of ordinary differential equations
* Using a random forest
* Using TensorFlow with Keras

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# R Machine Learning Projects

Course Snapshot

* **Course:** R Machine Learning Projects
* **Duration:** 2 days
* **Skill-level**: Foundation-level R Machine Learning Projects skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Master a range of machine learning domains with real-world projects using TensorFlow for R, H2O, MXNet, and more.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

R is one of the most popular languages when it comes to performing computational statistics (statistical computing) easily and exploring the mathematical side of machine learning. With this course, you will leverage the R ecosystem to build efficient machine learning applications that carry out intelligent tasks within your organization. This course will help you test your knowledge and skills, guiding you on how to build easily through to complex machine learning projects. You will first learn how to build powerful machine learning models with ensembles to predict employee attrition. Next, you’ll implement a joke recommendation engine and learn how to perform sentiment analysis on Amazon reviews. You’ll also explore different clustering techniques to segment customers using wholesale data. In addition to this, the course will get you acquainted with credit card fraud detection using autoencoders, and reinforcement learning to make predictions and win on a casino slot machine. By the end of the course, you will be equipped to confidently perform complex tasks to build research and commercial projects for automated operations.

Working in a hands-on learning environment, led by our R Machine Learning Projects expert instructor, students will learn about and explore:

* Master machine learning, deep learning, and predictive modeling concepts in R 3.5
* Build intelligent end-to-end projects for finance, retail, social media, and a variety of domains
* Implement smart cognitive models with helpful tips and best practices

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Explore deep neural networks and various frameworks that can be used in R
* Develop a joke recommendation engine to recommend jokes that match users’ tastes
* Create powerful ML models with ensembles to predict employee attrition
* Build autoencoders for credit card fraud detection
* Work with image recognition and convolutional neural networks
* Make predictions for casino slot machine using reinforcement learning
* Implement NLP techniques for sentiment analysis and customer segmentation

Audience & Pre-Requisites

This course is designed for developers wants to Master a range of machine learning domains with real-world projects using TensorFlow for R, H2O, MXNet, and more

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Exploring the Machine Learning Landscape**

* Exploring the Machine Learning Landscape
* ML versus software engineering
* Types of ML methods
* ML terminology – a quick review
* ML project pipeline
* Learning paradigm
* Datasets

1. **Predicting Employee Attrition Using Ensemble Models**

* Predicting Employee Attrition Using Ensemble Models
* Philosophy behind ensembling
* Getting started
* Understanding the attrition problem and the dataset
* K-nearest neighbors model for benchmarking the performance
* Bagging
* Randomization with random forests
* Boosting
* Stacking

1. **Implementing a Jokes Recommendation Engine**

* Implementing a Jokes Recommendation Engine
* Fundamental aspects of recommendation engines
* Getting started
* Understanding the Jokes recommendation problem and the dataset
* Building a recommendation system with an item-based collaborative filtering technique
* Building a recommendation system with a user-based collaborative filtering technique
* Building a recommendation system based on an association-rule mining technique
* Content-based recommendation engine
* Building a hybrid recommendation system for Jokes recommendations

1. **Sentiment Analysis of Amazon Reviews with NLP**

* Sentiment Analysis of Amazon Reviews with NLP
* The sentiment analysis problem
* Getting started
* Understanding the Amazon reviews dataset
* Building a text sentiment classifier with the BoW approach
* Understanding word embedding
* Building a text sentiment classifier with pretrained word2vec word embedding based on Reuters news corpus
* Building a text sentiment classifier with GloVe word embedding
* Building a text sentiment classifier with fastText

1. **Customer Segmentation Using Wholesale Data**

* Customer Segmentation Using Wholesale Data
* Understanding customer segmentation
* Understanding the wholesale customer dataset and the segmentation problem
* Identifying the customer segments in wholesale customer data using k-means clustering
* Identifying the customer segments in the wholesale customer data using DIANA
* Identifying the customer segments in the wholesale customers data using AGNES

1. **Image Recognition Using Deep Neural Networks**

* Image Recognition Using Deep Neural Networks
* Technical requirements
* Understanding computer vision
* Achieving computer vision with deep learning
* Introduction to the MXNet framework
* Understanding the MNIST dataset
* Implementing a deep learning network for handwritten digit recognition
* Implementing computer vision with pretrained models

1. **Credit Card Fraud Detection Using Autoencoders**

* Credit Card Fraud Detection Using Autoencoders
* Machine learning in credit card fraud detection
* Autoencoders explained
* The credit card fraud dataset
* Building AEs with the H2O library in R

1. **Automatic Prose Generation with Recurrent Neural Networks**

* Automatic Prose Generation with Recurrent Neural Networks
* Understanding language models
* Exploring recurrent neural networks
* Backpropagation through time
* Problems and solutions to gradients in RNN
* Building an automated prose generator with an RNN

1. **Winning the Casino Slot Machines with Reinforcement Learning**

* Winning the Casino Slot Machines with Reinforcement Learning
* Understanding RL
* Multi-arm bandit – real-world use cases
* Solving the MABP with UCB and Thompson sampling algorithms

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Large Scale Machine Learning with Python

Course Snapshot

* **Course:** Large Scale Machine Learning with Python
* **Duration:** 2 days
* **Skill-level**: Foundation-level R Large Scale Machine Learning with Python for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Learn to build powerful machine learning models quickly and deploy large-scale predictive applications
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Large Python machine learning projects involve new problems associated with specialized machine learning architectures and designs that many data scientists have yet to tackle. But finding algorithms and designing and building platforms that deal with large sets of data is a growing need. Data scientists have to manage and maintain increasingly complex data projects, and with the rise of big data comes an increasing demand for computational and algorithmic efficiency. Large Scale Machine Learning with Python uncovers a new wave of machine learning algorithms that meet scalability demands together with a high predictive accuracy. Dive into scalable machine learning and the three forms of scalability. Speed up algorithms that can be used on a desktop computer with tips on parallelization and memory allocation. Get to grips with new algorithms that are specifically designed for large projects and can handle bigger files, and learn about machine learning in big data environments. We will also cover the most effective machine learning techniques on a map reduce framework in Hadoop and Spark in Python.

Working in a hands-on learning environment, led by our Machine Learning with Python expert instructor, students will learn about and explore:

* Design, engineer and deploy scalable machine learning solutions with the power of Python
* Take command of Hadoop and Spark with Python for effective machine learning on a map reduce framework
* Build state-of-the-art models and develop personalized recommendations to perform machine learning at scale

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Apply the most scalable machine learning algorithms
* Work with modern state-of-the-art large-scale machine learning techniques
* Increase predictive accuracy with deep learning and scalable data-handling techniques
* Improve your work by combining the MapReduce framework with Spark
* Build powerful ensembles at scale
* Use data streams to train linear and non-linear predictive models from extremely large datasets using a single machine

Audience & Pre-Requisites

This course is designed for developers wants to Learn to build powerful machine learning models quickly and deploy large-scale predictive applications

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **First Steps to Scalability**

* First Steps to Scalability
* Explaining scalability in detail
* Python for large scale machine learning
* Python packages

1. **Scalable Learning in Scikit-learn**

* Scalable Learning in Scikit-learn
* Out-of-core learning
* Streaming data from sources
* Stochastic learning
* Feature management with data streams

1. **Fast SVM Implementations**

* Fast SVM Implementations
* Datasets to experiment with on your own
* Support Vector Machines
* Feature selection by regularization
* Including non-linearity in SGD
* Hyperparameter tuning

1. **Neural Networks and Deep Learning**

* Neural Networks and Deep Learning
* The neural network architecture
* Neural networks and regularization
* Neural networks and hyperparameter optimization
* Neural networks and decision boundaries
* Deep learning at scale with H2O
* Deep learning and unsupervised pretraining
* Deep learning with theanets
* Autoencoders and unsupervised learning

1. **Deep Learning with TensorFlow**

* Deep Learning with TensorFlow
* TensorFlow installation
* Machine learning on TensorFlow with SkFlow
* Keras and TensorFlow installation
* Convolutional Neural Networks in TensorFlow through Keras
* CNN's with an incremental approach
* GPU Computing

1. **Classification and Regression Trees at Scale**

* Classification and Regression Trees at Scale
* Bootstrap aggregation
* Random forest and extremely randomized forest
* Fast parameter optimization with randomized search
* CART and boosting
* XGBoost
* Out-of-core CART with H2O

1. **Unsupervised Learning at Scale**

* Unsupervised Learning at Scale
* Unsupervised methods
* Feature decomposition – PCA
* PCA with H2O
* Clustering – K-means
* K-means with H2O
* LDA

1. **Distributed Environments – Hadoop and Spark**

* Distributed Environments – Hadoop and Spark
* From a standalone machine to a bunch of nodes
* Setting up the VM
* The Hadoop ecosystem
* Spark

1. **Practical Machine Learning with Spark**

* Practical Machine Learning with Spark
* Setting up the VM for this chapter
* Sharing variables across cluster nodes
* Data preprocessing in Spark
* Machine learning with Spark
* Summary

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Exploratory Analysis with pandas

Course Snapshot

* **Course:** Exploratory Analysis with pandas
* **Duration:** 1 days
* **Skill-level**: Foundation-level Analysis with pandas skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Explore Python frameworks like pandas, Jupyter notebooks, and Matplotlib to build data pipelines and data visualization.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The pandas is a Python library that lets you manipulate, transform, and analyze data. It is a popular framework for exploratory data visualization and analyzing datasets and data pipelines based on their properties. This course will be your practical guide to exploring datasets using pandas. You will start by setting up Python, pandas, and Jupyter Notebooks. You will learn how to use Jupyter Notebooks to run Python code. We then show you how to get data into pandas and do some exploratory analysis, before learning how to manipulate and reshape data using pandas methods. You will also learn how to deal with missing data from your datasets, how to draw charts and plots using pandas and Matplotlib, and how to create some effective visualizations for your audience. Finally, you will wrapup your newly gained pandas knowledge by learning how to import data out of pandas into some popular file formats. By the end of this course, you will have a better understanding of exploratory analysis and how to build exploratory data pipelines with Python.

Working in a hands-on learning environment, led by our Analysis with pandas expert instructor, students will learn about and explore:

* Learn to set up data analysis pipelines with pandas and Jupyter notebooks
* Effective techniques for data selection, manipulation, and visualization
* Introduction to Matplotlib for interactive data visualization using charts and plots

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn how to read different kinds of data into pandas DataFrames for data analysis
* Manipulate, transform, and apply formulas to data imported into pandas DataFrames
* Use pandas to analyze and visualize different kinds of data to gain real-world insights
* Extract transformed data form pandas DataFrames and convert it into the formats your application expects
* Manipulate model time-series data, perform algorithmic trading, derive results on fixed and moving windows, and more

Audience & Pre-Requisites

This course is designed for developers wants to Explore Python frameworks like pandas, Jupyter notebooks, and Matplotlib to build data pipelines and data visualization

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Working with Different Kinds of Datasets**

* Working with Different Kinds of Datasets
* Using advanced options while reading data from CSV files
* Reading data from Excel files
* Handling missing data while reading
* Reading data from other popular formats

1. **Data Selection**

* Data Selection
* Introduction to datasets
* Selecting data from the dataset
* Sorting a pandas DataFrame
* Filtering rows of a pandas DataFrame
* Applying multiple filter criteria to a pandas DataFrame
* Using the axis parameter in pandas
* Using string methods in pandas
* Changing the datatype of a pandas series

1. **Manipulating, Transforming, and Reshaping Data**

* Manipulating, Transforming, and Reshaping Data
* Modifying a pandas DataFrame using the inplace parameter
* Using the groupby method
* Handling missing values in pandas
* Indexing in pandas DataFrames
* Renaming columns in a pandas DataFrame
* Removing columns from a pandas DataFrame
* Working with date and time series data
* Handling SettingWithCopyWarning
* Applying a function to a pandas series or DataFrame
* Merging and concatenating multiple DataFrames into one

1. **Visualizing Data Like a Pro**

* Visualizing Data Like a Pro
* Controlling plot aesthetics
* Choosing the colors for plots
* Plotting categorical data
* Plotting with Data-Aware Grids

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Integrating Hadoop

Course Snapshot

* **Course:** Integrating Hadoop
* **Duration:** 3 days
* **Skill-level**: Foundation-level Hadoop skills for Intermediate skilled team members. This is not a basic class.
* Targeted Audience: This course is geared for those who wants to a simple approach to harnessing the data.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In today’s time, data with value is branched off into numerous databases across multiple companies. The challenge is bringing the data together. Integrating Hadoop shows how Hadoop is used to collect and load the data on physical devices and the cloud. The book begins with an introduction of Hadoop and the types of data fit for it. Next, it focuses on assembling the integration team and gives an overview of workloads in the organization. You will also identify data sources for Hadoop, such as No SQL Databases and Legacy/Relational Databases, distinguish between ETL and ELT, and learn how to load and unload data into Hadoop. You will also practice managing big data using methods such as Upserts and Use HBase, and discover the advantages of real-time computing and the basic structure of streaming data architecture. Finally, you will interact with the master data of an organization and learn the top 10 mistakes people commit while integrating Hadoop data and how to avoid them.

Working in a hands-on learning environment, led by our Hadoop expert instructor, students will learn about and explore:

* Organize a successful Hadoop rollout
* Load, unload, and manage data in Hadoop
* Integrate Hadoop with the existing information infrastructure

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Study the different roles and responsibilities of the integration team
* Move data from one place to another with ETL and ELT
* Load the data into Hadoop using the original method, called Batch
* Find out how and where to use real-time computing framework Spark
* Discover project Apache Kafka and its role in streaming data processor
* Avoid common mistakes of integrating Hadoop data

Audience & Pre-Requisites

This course is designed for developers wants a simple approach to harnessing the data

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **1 Hadoop in Support of an Information Strategy**

* Introducing Hadoop
* Hadoop Distributions

1. **Preparing for Integration**

* Assembling the Integration Team
* Overview of Workloads for Hadoop in the Organization
* Identifying Data Sources for Hadoop
* Data Profiling
* Analyzing and Profiling Source Systems and Data

1. **ETL versus ELT**

* Continued Need for More Speed
* Preference with Hadoop
* Is ETL Dead?

1. **Loading Data into Hadoop**

* Advantages of Data Integration Tools
* Methods of Data Loading
* Path to Production
* How-To with Talend Big Data

1. **Managing Big Data**

* Big Data ELT
* Importance of Data Quality in Hadoop
* Stewardship of Big Data

1. **Unloading/Distributing Data from Hadoop**

* Hadoop Extracts
* Hadoop and SOA

1. **Apache Spark Cluster** Computing with Hadoop

* Advantages of Real-Time Computing
* How and Where to Use Spark

1. **Streaming Data**

* 8 Streaming Data
* Streaming Data Technology Distinctions

1. **Master Data Management and Big Data**

* Hadoop and Master Data Management
* Integrating with Master Data
* Data Virtualization
* MDM and Hadoop Disconnects

1. **Top 10 Mistakes Integrating Hadoop Data**

* 1. Integrating Data Without a Business Purpose
* 2. Integrating Data into Hadoop for an Enterprise Data Repository
* 3. Overemphasis on Data Integration Performance to the Detriment of Query Performance for Data Usage
* 4. Not Refining Data to the Point of Usefulness
* 5. Improper Node Specification
* 6. Over-Reliance on Open Source Hadoop
* 7. ETL instead of ELT
* 8. Using MapReduce to Load Hadoop
* 9. Using Spark through Hive to Load Hadoop
* 10. Ignoring the Quality of the Data Being Loaded

1. **Case Studies and Trends**

* Case Studies in Big Data Integration
* Trends in Hadoop and Summary of Ideas

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning Solutions

Course Snapshot

* **Course:** Machine Learning Solutions
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants Practical, hands-on solutions in Python to overcome any problem in Machine Learning.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning (ML) helps you find hidden insights from your data without the need for explicit programming. This course is your key to solving any kind of ML problem you might come across in your job. You’ll encounter a set of simple to complex problems while building ML models, and you'll not only resolve these problems, but you’ll also learn how to build projects based on each problem, with a practical approach and easy-to-follow examples. The course includes a wide range of applications: from analytics and NLP, to computer vision domains. Some of the applications you will be working on include stock price prediction, a recommendation engine, building a chat-bot, a facial expression recognition system, and many more. The problem examples we cover include identifying the right algorithm for your dataset and use cases, creating and labeling datasets, getting enough clean data to carry out processing, identifying outliers, overftting datasets, hyperparameter tuning, and more. Here, you'll also learn to make more timely and accurate predictions. In addition, you'll deal with more advanced use cases, such as building a gaming bot, building an extractive summarization tool for medical documents, and you'll also tackle the problems faced while building an ML model. By the end of this course, you'll be able to fine-tune your models as per your needs to deliver maximum productivity.

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Master the advanced concepts, methodologies, and use cases of machine learning
* Build ML applications for analytics, NLP and computer vision domains
* Solve the most common problems in building machine learning models

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Select the right algorithm to derive the best solution in ML domains
* Perform predictive analysis effciently using ML algorithms
* Predict stock prices using the stock index value
* Perform customer analytics for an e-commerce platform
* Build recommendation engines for various domains
* Build NLP applications for the health domain
* Build language generation applications using different NLP techniques
* Build computer vision applications such as facial emotion recognition

Audience & Pre-Requisites

This course is designed for developers wants Practical, hands-on solutions in Python to overcome any problem in Machine Learning

**Pre-Requisites:** Students should have familiar with

* Basics of ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Credit Risk Modeling**

* Credit Risk Modeling
* Introducing the problem statement
* Understanding the dataset
* Feature engineering for the baseline model
* Selecting machine learning algorithms
* Training the baseline model
* Understanding the testing matrix
* Testing the baseline model
* Problems with the existing approach
* Optimizing the existing approach
* Implementing the revised approach
* Best approach

1. **Stock Market Price Prediction**

* Stock Market Price Prediction
* Introducing the problem statement
* Collecting the dataset
* Understanding the dataset
* Data preprocessing and data analysis
* Feature engineering
* Selecting the Machine Learning algorithm
* Training the baseline model
* Understanding the testing matrix
* Testing the baseline model
* Exploring problems with the existing approach
* Understanding the revised approach
* Implementing the revised approach
* The best approach

1. **Customer Analytics**

* Customer Analytics
* Introducing customer segmentation
* Understanding the datasets
* Building the baseline approach
* Building the revised approach
* The best approach
* Customer segmentation for various domains

1. **Recommendation Systems for E-Commerce**

* Recommendation Systems for E-Commerce
* Introducing the problem statement
* Understanding the datasets
* Building the baseline approach
* Building the revised approach
* The best approach

1. **Sentiment Analysis**

* Sentiment Analysis
* Introducing problem statements
* Understanding the dataset
* Building the training and testing datasets for the baseline model
* Feature engineering for the baseline model
* Selecting the machine learning algorithm
* Training the baseline model
* Understanding the testing matrix
* Testing the baseline model
* Problem with the existing approach
* How to optimize the existing approach
* Implementing the revised approach
* The best approach

1. **Job Recommendation Engine**

* Job Recommendation Engine
* Introducing the problem statement
* Understanding the datasets
* Building the baseline approach
* Building the revised approach
* The best approach

1. **Text Summarization**

* Text Summarization
* Understanding the basics of summarization
* Introducing the problem statement
* Understanding datasets
* Building the baseline approach
* Building the revised approach
* The best approach

1. **Developing Chatbots**

* Developing Chatbots
* Introducing the problem statement
* Understanding datasets
* Building the basic version of a chatbot
* Implementing the rule-based chatbot
* Testing the rule-based chatbot
* Problems with the existing approach
* Implementing the revised approach
* Testing the revised approach
* Problems with the revised approach
* The best approach
* Discussing the hybrid approach

1. **Building a Real-Time Object Recognition App**

* Building a Real-Time Object Recognition App
* Introducing the problem statement
* Understanding the dataset
* Transfer Learning
* Setting up the coding environment
* Features engineering for the baseline model
* Selecting the machine learning algorithm
* Building the baseline model
* Understanding the testing metrics
* Testing the baseline model
* Problem with existing approach
* How to optimize the existing approach
* Implementing the revised approach
* The best approach

1. **Face Recognition and Face Emotion Recognition**

* Face Recognition and Face Emotion Recognition
* Introducing the problem statement
* Setting up the coding environment
* Understanding the concepts of face recognition
* Approaches for implementing face recognition
* Understanding the dataset for face emotion recognition
* Understanding the concepts of face emotion recognition
* Building the face emotion recognition model
* Understanding the testing matrix
* Testing the model
* Problems with the existing approach
* How to optimize the existing approach
* The best approach

1. **Building Gaming Bot**

* Building Gaming Bot
* Introducing the problem statement
* Setting up the coding environment
* Understanding Reinforcement Learning (RL)
* Basic Atari gaming bot
* Implementing the basic version of the gaming bot
* Building the Space Invaders gaming bot
* Implementing the Space Invaders gaming bot
* Building the Pong gaming bot
* Implementing the Pong gaming bot
* Just for fun - implementing the Flappy Bird gaming bot

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Turning Text into Gold: Taxonomies and Textual Analytics

Course Snapshot

* **Course:** Turning Text into Gold: Taxonomies and Textual Analytics
* **Duration:** 3 days
* **Skill-level**: Foundation-level Taxonomies and Textual Analytics skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Leverage the various aspects of taxonomies to extract insights from raw text.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

With businesses operating round the clock, a large amount of data gets generated. This data can be efficiently converted into useful knowledge that can take your business to a higher level. This course introduces you to the concept of taxonomies and how they are used to simplify and understand the text. You'll explore how to use taxonomies for textual analytics. It begins with a quick history of taxonomies and their earliest usage. You’ll learn about the different types of taxonomies (recursive, networked, hierarchical, and so on. You'll also learn about ontologies and understand how the ontology becomes a bridge between the worlds of technology and business and commerce. The later lessons of the course show how to find the taxonomies that you need for successful textual analytics, update your taxonomies to include the constantly-changing language, and extract meaningful information from raw text using different tools, such as textual disambiguation, document fracturing, and so on. By the end of this course, you’ll be able to utilize the various aspects of taxonomies for efficient textual analysis.

Working in a hands-on learning environment, led by our Taxonomies and Textual Analytics expert instructor, students will learn about and explore:

* Get familiar with taxonomies, their types, and applications
* Explore the role that taxonomies play in textual analytics
* Understand how textual analysis is used in various industries like banking, hospitality, airline, and more

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the difference between text-based and transaction-based data processing
* Explore ontologies and their use in textual analytics
* Study the different types of taxonomies
* Explore the various ways to customize the taxonomy
* Discover how to keep a taxonomy updated with new and changed words
* Build your own taxonomy

Audience & Pre-Requisites

This course is designed for developers wants to leverage the various aspects of taxonomies to extract insights from raw text.

**Pre-Requisites:** Students should have familiar with

* Basics of ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Introduction**

* Introduction

1. **Brief History of Taxonomies**

* Brief History of Taxonomies
* Insufficiency of Structured Data
* Manual Processing
* Evolution of Textual Analytic Technology

1. **Simple Taxonomies**

* Simple Taxonomies
* Taxonomy Components
* Taxonomies and Language

1. **Complex Taxonomies**

* Complex Taxonomies
* Hierarchical Taxonomies
* Networked Taxonomies
* More Applications of Taxonomies

1. **Ontologies**

* Ontologies

1. **Obtaining Taxonomies**

* 5: Obtaining Taxonomies
* Curated Taxonomies
* Building Your Own Taxonomy
* Qualifying the Nouns

1. **Changing Taxonomies**

* Changing Taxonomies

1. **Taxonomies as Databases**

* Taxonomies as Databases
* MoveRemove Processing
* Taxonomy Customization
* Word Pairs
* Transporting the Taxonomy

1. **Taxonomies and Data** Models

* Taxonomies and Data Models
* Types of Textual Data
* Types of Textual Dat

1. **Textual Analytics**

* 10: Textual Analytics
* Document Fracturing
* Named Value Processing
* Supporting Processes

1. **Stage 1 Processing**

* Stage 1 Processing
* Basic Refinements
* Custom Variables
* Inline Contextualization
* Proximity Analysis and Resolution
* Stop Word Processing
* Associative Word Processing
* Homographic Resolution
* Alternate Spelling
* Acronym Resolution
* Stemming
* Date Normalization

1. **Stage 2 Processing**

* 12: Stage 2 Processing
* Sentiment Analysis
* Negativity Analysis
* Medical Records

1. **Banking Analytics**

* 13: Banking Analytics
* Publicly Available Banking Data
* Comments Collected
* Textual Disambiguation
* Secondary Inference Analysis
* Visualization
* Interpreting the Dashboard
* Considering a Single Bank

1. **Call Center Analytics**

* Call Center Analytics
* What the Call Center Hears
* Processing the Narrative
* Examining the Dashboard
* Getting to Visualization

1. **Hospitality Analytics**

* Hospitality Analytics
* Voice of the Customer
* Analyzing Restaurant Feedback

1. **Airline Analytics**

* Airline Analytics

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Keras 2.x Projects

Course Snapshot

* **Course:** Keras 2.x Projects
* **Duration:** 3 days
* **Skill-level**: Foundation-level Keras 2.x Projects skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Demonstrate fundamentals of Deep Learning and neural network methodologies using Keras 2.x.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Keras 2.x Projects explains how to leverage the power of Keras to build and train state-of-the-art deep learning models through a series of practical projects that look at a range of real-world application areas. To begin with, you will quickly set up a deep learning environment by installing the Keras library. Through each of the projects, you will explore and learn the advanced concepts of deep learning and will learn how to compute and run your deep learning models using the advanced offerings of Keras. You will train fully-connected multilayer networks, convolutional neural networks, recurrent neural networks, autoencoders and generative adversarial networks using real-world training datasets. The projects you will undertake are all based on real-world scenarios of all complexity levels, covering topics such as language recognition, stock volatility, energy consumption prediction, faster object classification for self-driving vehicles, and more. By the end of this course, you will be well versed with deep learning and its implementation with Keras. You will have all the knowledge you need to train your own deep learning models to solve different kinds of problems.

Working in a hands-on learning environment, led by our Keras 2.x Projects expert instructor, students will learn about and explore:

* Experimental projects showcasing the implementation of high-performance deep learning models with Keras.
* Use-cases across reinforcement learning, natural language processing, GANs and computer vision.
* Build strong fundamentals of Keras in the area of deep learning and artificial intelligence

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Apply regression methods to your data and understand how the regression algorithm works
* Understand the basic concepts of classification methods and how to implement them in the Keras environment
* Import and organize data for neural network classification analysis
* Learn about the role of rectified linear units in the Keras network architecture
* Implement a recurrent neural network to classify the sentiment of sentences from movie reviews
* Set the embedding layer and the tensor sizes of a network

Audience & Pre-Requisites

This course is designed for developers wants to Demonstrate fundamentals of Deep Learning and neural network methodologies using Keras 2.x

**Pre-Requisites:** Students should have familiar with

* Basics of Keras
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Getting Started with Keras**

* Getting Started with Keras
* Introduction to Keras
* Keras backend options
* Installation
* Model fitting in Keras

1. **Modeling Real Estate Using** **Regression Analysis**

* Modeling Real Estate Using Regression Analysis
* Defining a regression problem
* Creating a linear regression model
* Multiple linear regression concepts
* Neural networks for regression using Keras

1. **Heart Disease Classification with Neural Networks**

* Heart Disease Classification with Neural Networks
* Basics of classification problems
* Different types of classification
* Pattern recognition using a Keras neural network

1. **Concrete Quality Prediction Using Deep Neural Networks**

* Concrete Quality Prediction Using Deep Neural Networks
* Basic concepts of ANNs
* Multilayer neural networks
* Implementing multilayer neural networks in Keras
* Building a Keras deep neural network model
* Improving the model performance by removing outliers

1. **Fashion Article Recognition Using Convolutional Neural Networks**

* Fashion Article Recognition Using Convolutional Neural Networks
* Understanding computer vision concepts
* Convolutional neural networks
* Common CNN architecture
* Implementing a CNN for object recognition

1. **Movie Reviews Sentiment Analysis Using Recurrent Neural Networks**

* Movie Reviews Sentiment Analysis Using Recurrent Neural Networks
* Sentiment analysis basic concepts
* Recurrent neural networks
* Classifying sentiment in movie reviews using an RNN

1. **Stock Volatility Forecasting Using Long Short-Term Memory**

* Stock Volatility Forecasting Using Long Short-Term Memory
* The basics of forecasting
* Time series analysis
* Time series models
* Long short-term memory in Keras
* Implementing an LSTM to forecast stock volatility

1. **Reconstruction of Handwritten Digit Images Using Autoencoders**

* Reconstruction of Handwritten Digit Images Using Autoencoders
* Basic concepts of image recognition
* Optical character recognition
* Generative neural networks
* The Keras autoencoders model
* Implementing autoencoder Keras layers to reconstruct handwritten digit images

1. **Robot Control System Using Deep Reinforcement Learning**

* Robot Control System Using Deep Reinforcement Learning
* Robot control overview
* The environment for controlling robot mobility
* Reinforcement learning basics
* Keras DQNs
* DQN to control a robot's mobility

1. **Reuters Newswire Topics Classifier in Keras**

* Reuters Newswire Topics Classifier in Keras
* Natural language processing
* The Natural Language Toolkit
* Implementing a DNN to label sentences

1. **What is Next?**

* What is Next?
* Deep learning methods
* Automated machine learning
* Differentiable neural computer
* Genetic programming and evolutionary strategies
* Inverse reinforcement learning

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Ensemble Machine Learning Cookbook

Course Snapshot

* **Course:** Ensemble Machine Learning Cookbook
* **Duration:** 3 days
* **Skill-level**: Foundation-level Ensemble Machine Learning Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Implement machine learning algorithms to build ensemble models using Keras, H2O, Scikit-Learn, Pandas and more.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Ensemble modeling is an approach used to improve the performance of machine learning models. It combines two or more similar or dissimilar machine learning algorithms to deliver superior intellectual powers. This course will help you to implement popular machine learning algorithms to cover different paradigms of ensemble machine learning such as boosting, bagging, and stacking. The Ensemble Machine Learning Cookbook will start by getting you acquainted with the basics of ensemble techniques and exploratory data analysis. You'll then learn to implement tasks related to statistical and machine learning algorithms to understand the ensemble of multiple heterogeneous algorithms. It will also ensure that you don't miss out on key topics, such as like resampling methods. As you progress, you’ll get a better understanding of bagging, boosting, stacking, and working with the Random Forest algorithm using real-world examples. The course will highlight how these ensemble methods use multiple models to improve machine learning results, as compared to a single model. In the concluding lessons, you'll delve into advanced ensemble models using neural networks, natural language processing, and more. You’ll also be able to implement models such as fraud detection, text categorization, and sentiment analysis. By the end of this course, you'll be able to harness ensemble techniques and the working mechanisms of machine learning algorithms to build intelligent models using individual recipes.

Working in a hands-on learning environment, led by our Ensemble Machine Learning Cookbook expert instructor, students will learn about and explore:

* Apply popular machine learning algorithms using a recipe-based approach
* Implement boosting, bagging, and stacking ensemble methods to improve machine learning models
* Discover real-world ensemble applications and encounter complex challenges in Kaggle competitions

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand how to use machine learning algorithms for regression and classification problems
* Implement ensemble techniques such as averaging, weighted averaging, and max-voting
* Get to grips with advanced ensemble methods, such as bootstrapping, bagging, and stacking
* Use Random Forest for tasks such as classification and regression
* Implement an ensemble of homogeneous and heterogeneous machine learning algorithms
* Learn and implement various boosting techniques, such as AdaBoost, Gradient Boosting Machine, and XGBoost

Audience & Pre-Requisites

This course is designed for developers wants to Implement machine learning algorithms to build ensemble models using Keras, H2O, Scikit-Learn, Pandas and more

**Pre-Requisites:** Students should have familiar with

* Basics of Python and ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Get Closer to Your Data**

* Get Closer to Your Data
* Introduction
* Data manipulation with Python
* Analyzing, visualizing, and treating missing values
* Exploratory data analysis

1. **Getting Started with Ensemble Machine Learning**

* Getting Started with Ensemble Machine Learning
* Introduction to ensemble machine learning
* Max-voting
* Averaging
* Weighted averaging

1. **Resampling Methods**

* Resampling Methods
* Introduction to sampling
* k-fold and leave-one-out cross-validation
* Bootstrapping

1. **Statistical and Machine Learning Algorithms**

* Statistical and Machine Learning Algorithms
* Technical requirements
* Multiple linear regression
* Logistic regression
* Naive Bayes
* Decision trees
* Support vector machines

1. **Bag the Models with Bagging**

* Bag the Models with Bagging
* Introduction
* Bootstrap aggregation
* Ensemble meta-estimators
* Bagging regressors

1. **When in Doubt, Use Random Forests**

* When in Doubt, Use Random Forests
* Introduction to random forests
* Implementing a random forest for predicting credit card defaults using scikit-learn
* Implementing random forest for predicting credit card defaults using H2O

1. **Boosting Model Performance with Boosting**

* Boosting Model Performance with Boosting
* Introduction to boosting
* Implementing AdaBoost for disease risk prediction using scikit-learn
* Implementing a gradient boosting machine for disease risk prediction using scikit-learn
* Implementing the extreme gradient boosting method for glass identification using XGBoost with scikit-learn

1. **Blend It with Stacking**

* Blend It with Stacking
* Technical requirements
* Understanding stacked generalization
* Implementing stacked generalization by combining predictions
* Implementing stacked generalization for campaign outcome prediction using H2O

1. **Homogeneous Ensembles Using Keras**

* Homogeneous Ensembles Using Keras
* Introduction
* An ensemble of homogeneous models for energy prediction
* An ensemble of homogeneous models for handwritten digit classification

1. **Heterogeneous Ensemble Classifiers Using H2O**

* Heterogeneous Ensemble Classifiers Using H2O
* Introduction
* Predicting credit card defaulters using heterogeneous ensemble classifiers

1. **Heterogeneous Ensemble for Text Classification Using NLP**

* Heterogeneous Ensemble for Text Classification Using NLP
* Introduction
* Spam filtering using an ensemble of heterogeneous algorithms
* Sentiment analysis of movie reviews using an ensemble model

1. **Homogenous Ensemble for Multiclass Classification Using Keras**

* Homogenous Ensemble for Multiclass Classification Using Keras
* Introduction
* An ensemble of homogeneous models to classify fashion products

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Intelligent Mobile Projects with TensorFlow

Course Snapshot

* **Course:** Intelligent Mobile Projects with TensorFlow
* **Duration:** 3 days
* **Skill-level**: Foundation-level Intelligent Mobile Projects with TensorFlow skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Create Deep Learning and Reinforcement Learning apps for multiple platforms with TensorFlow.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

As a developer, you always need to keep an eye out and be ready for what will be trending soon, while also focusing on what's trending currently. So, what's better than learning about the integration of the best of both worlds, the present and the future? Artificial Intelligence (AI) is widely regarded as the next big thing after mobile, and Google's TensorFlow is the leading open source machine learning framework, the hottest branch of AI. This course covers more than 10 complete iOS, Android, and Raspberry Pi apps powered by TensorFlow and built from scratch, running all kinds of cool TensorFlow models offline on-device: from computer vision, speech and language processing to generative adversarial networks and AlphaZero-like deep reinforcement learning. You’ll learn how to use or retrain existing TensorFlow models, build your own models, and develop intelligent mobile apps running those TensorFlow models. You'll learn how to quickly build such apps with step-by-step tutorials and how to avoid many pitfalls in the process with lots of hard-earned troubleshooting tips..

Working in a hands-on learning environment, led by our Intelligent Mobile Projects with TensorFlow expert instructor, students will learn about and explore:

* Build TensorFlow-powered AI applications for mobile and embedded devices
* Learn modern AI topics such as computer vision, NLP, and deep reinforcement learning
* Get practical insights and exclusive working code not available in the TensorFlow documentation

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Classify images with transfer learning
* Detect objects and their locations
* Transform pictures with amazing art styles
* Understand simple speech commands
* Describe images in natural language
* Recognize drawing with Convolutional Neural Network and Long Short-Term Memory
* Predict stock price with Recurrent Neural Network in TensorFlow and Keras
* Generate and enhance images with generative adversarial networks
* Build AlphaZero-like mobile game app in TensorFlow and Keras
* Use TensorFlow Lite and Core ML on mobile
* Develop TensorFlow apps on Raspberry Pi that can move, see, listen, speak, and learn

Audience & Pre-Requisites

This course is designed for developers wants to Create Deep Learning and Reinforcement Learning apps for multiple platforms with TensorFlow

**Pre-Requisites:** Students should have familiar with

* Basics of Python and ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Getting Started with Mobile TensorFlow**

* Getting Started with Mobile TensorFlow
* Setting up TensorFlow
* Setting up Xcode
* Setting up Android Studio
* TensorFlow Mobile vs TensorFlow Lite
* Running sample TensorFlow iOS apps
* Running sample TensorFlow Android apps

1. **Classifying Images with Transfer Learning**

* Classifying Images with Transfer Learning
* Transfer learning – what and why
* Retraining using the Inception v3 model
* Retraining using MobileNet models
* Using the retrained models in the sample iOS app
* Using the retrained models in the sample Android app
* Adding TensorFlow to your own iOS app
* Adding TensorFlow to your own Android app

1. **Detecting Objects and Their Locations**

* Detecting Objects and Their Locations
* Object detection–a quick overview
* Setting up the TensorFlow Object Detection API
* Retraining SSD-MobileNet and Faster RCNN models
* Using object detection models in iOS
* Using YOLO2–another object-detection model

1. **Transforming Pictures with Amazing Art Styles**

* Transforming Pictures with Amazing Art Styles
* Neural Style Transfer – a quick overview
* Training fast neural-style transfer models
* Using fast neural-style transfer models in iOS
* Using fast neural-style transfer models in Android
* Using the TensorFlow Magenta multi-style model in iOS
* Using the TensorFlow Magenta multi-style model in Android

1. **Understanding Simple Speech Commands**

* Understanding Simple Speech Commands
* Speech recognition – a quick overview
* Training a simple commands recognition model
* Using a simple speech recognition model in Android
* Using a simple speech recognition model in iOS with Objective-C
* Using a simple speech recognition model in iOS with Swift

1. **Describing Images in Natural Language**

* Describing Images in Natural Language
* Image captioning – how it works
* Training and freezing an image captioning model
* Transforming and optimizing the image captioning model
* Using the image captioning model in iOS
* Using the image captioning model in Android

1. **Recognizing Drawing with CNN and LSTM**

* Recognizing Drawing with CNN and LSTM
* Drawing classification – how it works
* Training, predicting, and preparing the drawing classification model
* Using the drawing classification model in iOS
* Using the drawing classification model in Android

1. **Predicting Stock Price with RNN**

* Predicting Stock Price with RNN
* RNN and stock price prediction – what and how
* Using the TensorFlow RNN API for stock price prediction
* Using the Keras RNN LSTM API for stock price prediction
* Running the TensorFlow and Keras models on iOS
* Running the TensorFlow and Keras models on Android

1. **Generating and Enhancing Images with GAN**

* Generating and Enhancing Images with GAN
* GAN – what and why
* Building and training GAN models with TensorFlow
* Using the GAN models in iOS
* Using the GAN models in Android

1. **Building an AlphaZero-like Mobile Game App**

* Building an AlphaZero-like Mobile Game App
* AlphaZero – how does it work?
* Training and testing an AlphaZero-like model for Connect 4
* Using the model in iOS to play Connect 4
* Using the model in Android to play Connect 4

1. **Using TensorFlow Lite and Core ML on Mobile**

* Using TensorFlow Lite and Core ML on Mobile
* TensorFlow Lite – an overview
* Using TensorFlow Lite in iOS
* Using TensorFlow Lite in Android
* Core ML for iOS – an overview
* Using Core ML with Scikit-Learn machine learning
* Using Core ML with Keras and TensorFlow

1. **Developing TensorFlow Apps on Raspberry Pi**

* Developing TensorFlow Apps on Raspberry Pi
* Setting up Raspberry Pi and making it move
* Setting up TensorFlow on Raspberry Pi
* Image recognition and text to speech
* Audio recognition and robot movement
* Reinforcement learning on Raspberry Pi

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Jupyter Cookbook

Course Snapshot

* **Course:** Jupyter Cookbook
* **Duration:** 3 days
* **Skill-level**: Foundation-level Jupyter Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Leverage the power of the popular Jupyter notebooks to simplify your data science tasks without any hassle.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Jupyter has garnered a strong interest in the data science community of late, as it makes common data processing and analysis tasks much simpler. This course is for data science professionals who want to master various tasks related to Jupyter to create efficient, easy-to-share, scientific applications. The course starts with recipes on installing and running the Jupyter Notebook system on various platforms and configuring the various packages that can be used with it. You will then see how you can implement different programming languages and frameworks, such as Python, R, Julia, JavaScript, Scala, and Spark on your Jupyter Notebook. This course contains intuitive recipes on building interactive widgets to manipulate and visualize data in real time, sharing your code, creating a multi-user environment, and organizing your notebook. You will then get hands-on experience with Jupyter Labs, microservices, and deploying them on the web. By the end of this course, you will have taken your knowledge of Jupyter to the next level to perform all key tasks associated with it.

Working in a hands-on learning environment, led by our Jupyter Cookbook expert instructor, students will learn about and explore:

* Create and share interactive documents with live code, text and visualizations
* Integrate popular programming languages such as Python, R, Julia, Scala with Jupyter
* Develop your widgets and interactive dashboards with these innovative recipes

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Install Jupyter and configure engines for Python, R, Scala and more
* Access and retrieve data on Jupyter Notebooks
* Create interactive visualizations and dashboards for different scenarios
* Convert and share your dynamic codes using HTML, JavaScript, Docker, and more
* Create custom user data interactions using various Jupyter widgets
* Manage user authentication and file permissions
* Interact with Big Data to perform numerical computing and statistical modeling
* Get familiar with Jupyter's next-gen user interface - JupyterLab

Audience & Pre-Requisites

This course is designed for developers wants to Leverage the power of the popular Jupyter notebooks to simplify your data science tasks without any hassle

**Pre-Requisites:** Students should have familiar with

* Basics of Python and ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Installation and Setting up the Environment**

* Installation and Setting up the Environment
* Introduction
* Installing Jupyter on Windows
* Installing Jupyter on the Mac
* Installing Jupyter on Linux
* Installing Jupyter on a server

1. **Adding an Engine**

* Adding an Engine
* Introduction
* Adding the Python 3 engine
* Adding the R engine
* Adding the Julia engine
* Adding the JavaScript engine
* Adding the Scala engine
* Adding the Spark engine

1. **Accessing and Retrieving Data**

* Accessing and Retrieving Data
* Introduction
* Reading CSV files
* Reading JSON files
* Accessing a database
* Reading flat files
* Reading text files

1. **Visualizing Your Analytics**

* Visualizing Your Analytics
* Introduction
* Generating a line graph using Python
* Generating a histogram using Python
* Generating a density map using Python
* Plotting 3D data using Python
* Present a user-interactive graphic using Python
* Visualizing with R
* Generate a regression line of data using R
* Generate an R lowess line graph
* Producing a Scatter plot matrix using R
* Producing a bar chart using R
* Producing a word cloud using R
* Visualizing with Julia
* Drawing a Julia scatter diagram of Iris data using Gadfly
* Drawing a Julia histogram using Gadfly
* Drawing a Julia line graph using the Winston package

1. **Working with Widgets**

* Working with Widgets
* Introduction
* What are widgets?
* Using ipyleaflet widgets
* Using ipywidgets
* Using a widget container
* Using an interactive widget
* Using an interactive text widget
* Linking widgets together
* Another ipywidgets linking example
* Using a cookie cutter widget
* Developing an OPENGL widget
* Creating a simple orbit of one object
* Using a complex orbit of multiple objects

1. **Jupyter Dashboards**

* Jupyter Dashboards
* Introduction
* What is Jupyter dashboards?
* Creating an R dashboard
* Create a Python dashboard
* Creating a Julia dashboard
* Develop a JavaScript (Node.js) dashboard

1. **Sharing Your Code**

* Sharing Your Code
* Introduction
* Using a Notebook server
* Using a web server
* Sharing your Notebook through a public server
* Sharing your Notebook through Docker
* Sharing your Notebook using nbviewer
* Converting your Notebook into a different format
* Converting Notebooks to R
* Converting Notebooks to HTML
* Converting Notebooks to Markdown
* Converting Notebooks to reStructedText
* Converting Notebooks to Latex
* Converting Notebooks to PDF

1. **Multiuser Jupyter**

* Multiuser Jupyter
* Introduction
* Why multiuser?
* Providing multiuser with JupyterHub
* Providing multiuser with Docker
* Running your Notebook in Google Cloud Platform
* Running your Notebook in AWS
* Running your Notebook in Azure

1. **Interacting with Big Data**

* Interacting with Big Data
* Introduction
* Obtaining a word count from a big-text data source
* Obtaining a sorted word count from a big-text source
* Examining big-text log file access
* Computing prime numbers using parallel operations
* Analyzing big-text data
* Analyzing big data history files

1. **Jupyter Security**

* Jupyter Security
* Introduction
* Security mechanisms built into Jupyter
* Using SSL
* The Jupyter trust model
* Controlling network access
* Additional practices

1. **Jupyter Labs**

* Jupyter Labs
* Introduction
* Installing and starting JupyterLab
* JupyterLab display
* JupyterLab menus
* Starting a Notebook
* Starting a console

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning for Algorithmic Trading

Course Snapshot

* **Course:** Machine Learning for Algorithmic Trading
* **Duration:** 5 days
* **Skill-level**: Foundation-level Machine Learning for Algorithmic Tradingskills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Explore effective trading strategies in real-world markets using NumPy, spaCy, pandas, scikit-learn, and Keras.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The explosive growth of digital data has boosted the demand for expertise in trading strategies that use machine learning (ML). This course enables you to use a broad range of supervised and unsupervised algorithms to extract signals from a wide variety of data sources and create powerful investment strategies. This course shows how to access market, fundamental, and alternative data via API or web scraping and offers a framework to evaluate alternative data. You’ll practice the ML workﬂow from model design, loss metric definition, and parameter tuning to performance evaluation in a time series context. You will understand ML algorithms such as Bayesian and ensemble methods and manifold learning, and will know how to train and tune these models using pandas, statsmodels, sklearn, PyMC3, xgboost, lightgbm, and catboost. This course also teaches you how to extract features from text data using spaCy, classify news and assign sentiment scores, and to use gensim to model topics and learn word embeddings from financial reports. You will also build and evaluate neural networks, including RNNs and CNNs, using Keras and PyTorch to exploit unstructured data for sophisticated strategies. Finally, you will apply transfer learning to satellite images to predict economic activity and use reinforcement learning to build agents that learn to trade in the OpenAI Gym.

Working in a hands-on learning environment, led by our Machine Learning for Algorithmic Trading expert instructor, students will learn about and explore:

* Implement machine learning algorithms to build, train, and validate algorithmic models
* Create your own algorithmic design process to apply probabilistic machine learning approaches to trading decisions
* Develop neural networks for algorithmic trading to perform time series forecasting and smart analytics

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Implement machine learning techniques to solve investment and trading problems
* Leverage market, fundamental, and alternative data to research alpha factors
* Design and fine-tune supervised, unsupervised, and reinforcement learning models
* Optimize portfolio risk and performance using pandas, NumPy, and scikit-learn
* Integrate machine learning models into a live trading strategy on Quantopian
* Evaluate strategies using reliable backtesting methodologies for time series
* Design and evaluate deep neural networks using Keras, PyTorch, and TensorFlow
* Work with reinforcement learning for trading strategies in the OpenAI Gym

Audience & Pre-Requisites

This course is designed for developers wants to Explore effective trading strategies in real-world markets using NumPy, spaCy, pandas, scikit-learn, and Keras

**Pre-Requisites:** Students should have familiar with

* Basics of Python and ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Machine Learning for Trading**

* Machine Learning for Trading
* How to read this course
* The rise of ML in the investment industry
* Design and execution of a trading strategy
* ML and algorithmic trading strategies

1. **Market and Fundamental Data**

* Market and Fundamental Data
* How to work with market data
* How to work with fundamental data
* Efficient data storage with pandas

1. **Alternative Data for Finance**

* Alternative Data for Finance
* The alternative data revolution
* Evaluating alternative datasets
* The market for alternative data
* Working with alternative data

1. **Alpha Factor Research**

* Alpha Factor Research
* Engineering alpha factors
* Seeking signals – how to use zipline
* Separating signal and noise – how to use alphalens
* Alpha factor resources

1. **Strategy Evaluation**

* Strategy Evaluation
* How to build and test a portfolio with zipline
* How to measure performance with pyfolio
* How to avoid the pitfalls of backtesting
* How to manage portfolio risk and return

1. **The Machine Learning Process**

* The Machine Learning Process
* Learning from data
* The machine learning workflow

1. **Linear Models**

* Linear Models
* Linear regression for inference and prediction
* The multiple linear regression model
* How to build a linear factor model
* Shrinkage methods – regularization for linear regression
* How to use linear regression to predict returns
* Linear classification

1. **Time Series Models**

* Time Series Models
* Analytical tools for diagnostics and feature extraction
* Univariate time series models
* Multivariate time series models

1. **Bayesian Machine Learning**

* Bayesian Machine Learning
* How Bayesian machine learning works
* Probabilistic programming with PyMC3

1. **Decision Trees and Random Forests**

* Decision Trees and Random Forests
* Decision trees
* Random forests

1. **Gradient Boosting Machines**

* Gradient Boosting Machines
* Adaptive boosting
* Gradient boosting machines
* Fast scalable GBM implementations
* How to interpret GBM results

1. **Unsupervised Learning**

* Unsupervised Learning
* Dimensionality reduction
* Clustering

1. **Working with Text Data**

* Working with Text Data
* How to extract features from text data
* From text to tokens – the NLP pipeline
* From tokens to numbers – the document-term matrix
* Text classification and sentiment analysis

1. **Topic Modeling**

* Topic Modeling
* Learning latent topics: goals and approaches
* Latent semantic indexing
* Probabilistic latent semantic analysis
* Latent Dirichlet allocation

1. **Word Embeddings**

* Word Embeddings
* How word embeddings encode semantics
* Word vectors from SEC filings using gensim
* Sentiment analysis with Doc2vec
* Bonus – Word2vec for translation

1. **Deep Learning**

* Deep Learning
* Deep learning and AI
* How to design a neural network
* How to build a neural network using Python
* How to train a neural network
* How to use DL libraries
* How to optimize neural network architectures

1. **Convolutional Neural Networks**

* Convolutional Neural Networks
* How ConvNets work
* How to design and train a CNN using Python
* Transfer learning – faster training with less data
* How to detect objects
* Recent developments

1. **Recurrent Neural Networks**

* Recurrent Neural Networks
* How RNNs work
* How to build and train RNNs using Python

1. **Autoencoders and Generative Adversarial Nets**

* Autoencoders and Generative Adversarial Nets
* How autoencoders work
* Designing and training autoencoders using Python
* How GANs work

1. **Reinforcement Learning**

* Reinforcement Learning
* Key elements of RL
* How to solve RL problems
* Dynamic programming – Value and Policy iteration
* Q-learning
* Deep reinforcement learning
* Reinforcement learning for trading

1. **Next Steps**

* Next Steps
* Key takeaways and lessons learned
* ML for trading in practice
* Conclusion

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with the Elastic Stack

Course Snapshot

* **Course:** Machine Learning with the Elastic Stack
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning with the Elastic Stack skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Leverage Elastic Stack’s machine learning features to gain valuable insight from your data.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine Learning with the Elastic Stack is a comprehensive overview of the embedded commercial features of anomaly detection and forecasting. The course starts with installing and setting up Elastic Stack. You will perform time series analysis on varied kinds of data, such as log files, network flows, application metrics, and financial data. As you progress through the lessons, you will deploy machine learning within the Elastic Stack for logging, security, and metrics. In the concluding lessons, you will see how machine learning jobs can be automatically distributed and managed across the Elasticsearch cluster and made resilient to failure. By the end of this course, you will understand the performance aspects of incorporating machine learning within the Elastic ecosystem and create anomaly detection jobs and view results from Kibana directly.

Working in a hands-on learning environment, led by our Machine Learning with the Elastic Stack expert instructor, students will learn about and explore:

* Combine machine learning with the analytic capabilities of Elastic Stack
* Analyze large volumes of search data and gain actionable insight from them
* Use external analytical tools with your Elastic Stack to improve its performance

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Install the Elastic Stack to use machine learning features
* Understand how Elastic machine learning is used to detect a variety of anomaly types
* Apply effective anomaly detection to IT operations and security analytics
* Leverage the output of Elastic machine learning in custom views, dashboards, and proactive alerting
* Combine your created jobs to correlate anomalies of different layers of infrastructure
* Learn various tips and tricks to get the most out of Elastic machine learning

Audience & Pre-Requisites

This course is designed for developers wants Leverage Elastic Stack’s machine learning features to gain valuable insight from your data

**Pre-Requisites:** Students should have familiar with

* Basics of ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Machine Learning for IT**

* Machine Learning for IT
* Overcoming the historical challenges
* Theory of operation
* Operationalization
* Supporting indices
* The orchestration

1. **Installing the Elastic Stack with Machine Learning**

* Installing the Elastic Stack with Machine Learning
* Installing the Elastic Stack
* A guided tour of Elastic ML features

1. **Event Change Detection**

* Event Change Detection
* How to understand the normal rate of occurrence
* Exploring count functions
* Counting in population analysis
* Detecting things that rarely occur
* Counting message-based logs via categorization

1. **IT Operational Analytics and Root Cause Analysis**

* IT Operational Analytics and Root Cause Analysis
* Holistic application visibility
* Data organization
* Bringing it all together for root cause analysis

1. **Security Analytics with Elastic Machine Learning**

* Security Analytics with Elastic Machine Learning
* Security in the field
* Threat hunting architecture
* Investigation analytics

1. **Alerting on ML Analysis**

* Alerting on ML Analysis
* Results presentation
* The results index
* Alerts from the Machine Learning UI in Kibana
* Creating ML alerts manually

1. **Using Elastic ML Data in Kibana Dashboards**

* Using Elastic ML Data in Kibana Dashboards
* Visualization options in Kibana
* Preparing data for anomaly detection analysis
* Building the visualizations

1. **Using Elastic ML with Kibana Canvas**

* Using Elastic ML with Kibana Canvas
* Introduction to Canvas
* Building Elastic ML Canvas slides

1. **Forecasting**

* Forecasting
* Forecasting versus prophesying
* Forecasting use cases
* Forecasting – theory of operation
* Single time series forecasting
* Forecast results
* Multiple time series forecasting

1. **ML Tips and Tricks**

* ML Tips and Tricks
* Job groups
* Influencers in split versus non-split jobs
* Using ML on scripted fields
* Using one-sided ML functions to your advantage
* Ignoring time periods
* Don't over-engineer the use case
* ML job throughput considerations
* Top-down alerting by leveraging custom rules
* Sizing ML deployments

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Python Machine Learning Blueprints

Course Snapshot

* **Course:** Python Machine Learning Blueprints
* **Duration:** 3 days
* **Skill-level**: Foundation-level Python Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Discover a project-based approach to mastering machine learning concepts by applying them to everyday problems using libraries such as scikit-learn, TensorFlow, and Keras.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning is transforming the way we understand and interact with the world around us. This course is the perfect guide for you to put your knowledge and skills into practice and use the Python ecosystem to cover key domains in machine learning. This edition covers a range of libraries from the Python ecosystem, including TensorFlow and Keras, to help you implement real-world machine learning projects. The course begins by giving you an overview of machine learning with Python. With the help of complex datasets and optimized techniques, you’ll go on to understand how to apply advanced concepts and popular machine learning algorithms to real-world projects. Next, you’ll cover projects from domains such as predictive analytics to analyze the stock market and recommendation systems for GitHub repositories. In addition to this, you’ll also work on projects from the NLP domain to create a custom news feed using frameworks such as scikit-learn, TensorFlow, and Keras. Following this, you’ll learn how to build an advanced chatbot, and scale things up using PySpark. In the concluding lessons, you can look forward to exciting insights into deep learning and you'll even create an application using computer vision and neural networks. By the end of this course, you’ll be able to analyze data seamlessly and make a powerful impact through your projects.

Working in a hands-on learning environment, led by our Python Machine Learning Blueprints expert instructor, students will learn about and explore:

* Get to grips with Python's machine learning libraries including scikit-learn, TensorFlow, and Keras
* Implement advanced concepts and popular machine learning algorithms in real-world projects
* Build analytics, computer vision, and neural network projects

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the Python data science stack and commonly used algorithms
* Build a model to forecast the performance of an Initial Public Offering (IPO) over an initial discrete trading window
* Understand NLP concepts by creating a custom news feed
* Create applications that will recommend GitHub repositories based on ones you’ve starred, watched, or forked
* Gain the skills to build a chatbot from scratch using PySpark
* Develop a market-prediction app using stock data
* Delve into advanced concepts such as computer vision, neural networks, and deep learning

Audience & Pre-Requisites

This course is designed for developers wants to Discover a project-based approach to mastering machine learning concepts by applying them to everyday problems using libraries such as scikit-learn, TensorFlow, and Keras

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **The Python Machine Learning Ecosystem**

* The Python Machine Learning Ecosystem
* Data science/machine learning workflow
* Python libraries and functions for each stage of the data science workflow
* Setting up your machine learning environment

1. **Build an App to Find Underpriced Apartments**

* Build an App to Find Underpriced Apartments
* Sourcing apartment listing data
* Inspecting and preparing the data
* Visualizing our data
* Visualizing the data
* Modeling the data
* Extending the model

1. **Build an App to Find Cheap Airfares**

* Build an App to Find Cheap Airfares
* Sourcing airfare pricing data
* Retrieving fare data with advanced web scraping
* Parsing the DOM to extract pricing data
* Identifying outlier fares with anomaly detection techniques
* Sending real-time alerts using IFTTT
* Putting it all together

1. **Forecast the IPO Market Using Logistic Regression**

* Forecast the IPO Market Using Logistic Regression
* The IPO market
* Data cleansing and feature engineering
* Binary classification with logistic regression
* Generating the importance of a feature from our model

1. **Create a Custom Newsfeed**

* Create a Custom Newsfeed
* Creating a supervised training set with Pocket
* Using the Embedly API to download story bodies
* Basics of Natural Language Processing
* Support Vector Machines
* IFTTT integration with feeds, Google Sheets, and email
* Setting up your daily personal newsletter

1. **Predict whether Your Content Will Go Viral**

* Predict whether Your Content Will Go Viral
* What does research tell us about virality?
* Sourcing shared counts and content
* Exploring the features of shareability
* Building a predictive content scoring model

1. **Use Machine Learning to Forecast the Stock Market**

* Use Machine Learning to Forecast the Stock Market
* Types of market analysis
* What does research tell us about the stock market?
* How to develop a trading strategy
* Building the regression model

1. **Classifying Images with** **Convolutional Neural Networks**

* Classifying Images with Convolutional Neural Networks
* Image-feature extraction
* Convolutional neural networks
* Building a convolutional neural network to classify images in the Zalando Research dataset, using Keras

1. **Building a Chatbot**

* Building a Chatbot
* The Turing Test
* The history of chatbots
* The design of chatbots
* Building a chatbot
* Sequence-to-sequence modeling for chatbots

1. **Build a Recommendation Engine**

* Build a Recommendation Engine
* Collaborative filtering
* Content-based filtering
* Hybrid systems
* Building a recommendation engine

1. **What's Next?**

* What's Next?
* Summary of the projects

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning for Mobile

Course Snapshot

* **Course:** Machine Learning for Mobile
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning for Mobile skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Leverage the power of machine learning on mobiles and build intelligent mobile applications with ease.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning presents an entirely unique opportunity in software development. It allows smartphones to produce an enormous amount of useful data that can be mined, analyzed, and used to make predictions. This course will help you master machine learning for mobile devices with easy-to-follow, practical examples. You will begin with an introduction to machine learning on mobiles and grasp the fundamentals so you become well-acquainted with the subject. You will master supervised and unsupervised learning algorithms, and then learn how to build a machine learning model using mobile-based libraries such as Core ML, TensorFlow Lite, ML Kit, and Fritz on Android and iOS platforms. In doing so, you will also tackle some common and not-so-common machine learning problems with regard to Computer Vision and other real-world domains. By the end of this course, you will have explored machine learning in depth and implemented on-device machine learning with ease, thereby gaining a thorough understanding of how to run, create, and build real-time machine-learning applications on your mobile devices.

Working in a hands-on learning environment, led by our Machine Learning for Mobile expert instructor, students will learn about and explore:

* Build smart mobile applications for Android and iOS devices
* Use popular machine learning toolkits such as Core ML and TensorFlow Lite
* Explore cloud services for machine learning that can be used in mobile apps

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Build intelligent machine learning models that run on Android and iOS
* Use machine learning toolkits such as Core ML, TensorFlow Lite, and more
* Learn how to use Google Mobile Vision in your mobile apps
* Build a spam message detection system using Linear SVM
* Using Core ML to implement a regression model for iOS devices
* Build image classification systems using TensorFlow Lite and Core ML

Audience & Pre-Requisites

This course is designed for developers wants to Leverage the power of machine learning on mobiles and build intelligent mobile applications with ease

**Pre-Requisites:** Students should have familiar with

* Basics of ML
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Introduction to Machine Learning on Mobile**

* Introduction to Machine Learning on Mobile
* Definition of machine learning
* The machine learning process
* Types of learning
* Why use machine learning on mobile devices?

1. **Supervised and Unsupervised Learning Algorithms**

* Supervised and Unsupervised Learning Algorithms
* Introduction to supervised learning algorithms
* Deep dive into supervised learning algorithms
* Introduction to unsupervised learning algorithms
* Deep dive into unsupervised learning algorithms

1. **Random Forest on iOS**

* Random Forest on iOS
* Introduction to algorithms
* Solving the problem using random forest in Core ML

1. **TensorFlow Mobile in Android**

* TensorFlow Mobile in Android
* An introduction to TensorFlow
* The architecture of a mobile machine learning application
* Writing the mobile application using the TensorFlow model

1. **Regression Using Core ML in iOS**

* Regression Using Core ML in iOS
* Introduction to regression
* Understanding the basics of Core ML
* Solving the problem using regression in Core ML

1. **The ML Kit SDK**

* The ML Kit SDK
* Understanding ML Kit
* Creating a text recognition app using Firebase on-device APIs
* Creating a text recognition app using Firebase on-cloud APIs
* Face detection using ML Kit

1. **Spam Message Detection**

* Spam Message Detection
* Understanding NLP
* Understanding linear SVM algorithm
* Solving the problem using linear SVM in Core ML

1. **Fritz**

* Fritz
* Introduction to Fritz
* Hand-on samples using Fritz

1. **Neural Networks on Mobile**

* Neural Networks on Mobile
* Introduction to neural networks
* Image recognition solution
* Creating a TensorFlow image recognition model
* Handwritten digit recognition solution
* Introduction to Keras
* Installing Keras
* Solving the problem

1. **Mobile Application Using Google Vision**

* Mobile Application Using Google Vision
* Features of Google Cloud Vision
* Sample mobile application using Google Cloud Vision

1. **The Future of ML on Mobile Applications**

* The Future of ML on Mobile Applications
* Key ML mobile applications
* Key innovation areas
* Opportunities for stakeholders

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Analysis and Business Modeling with Excel 2013

**Course Snapshot**

* **Course:** Excel Data Analysis
* **Duration:** 3 days
* **Skill-level**: Foundation-level Excel Data Analysis skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Manage, analyze, and visualize data with Microsoft Excel 2013 to transform raw data into ready to use information
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Data Analysis and Business Modeling with Excel 2013 is one of the easiest to use data analysis tools you will ever come across. Its simplicity and powerful features has made it the go to tool for all your data needs. Complex operations with Excel, such as creating charts and graphs, visualization, and analyzing data make it a great tool for managers, data scientists, financial data analysts, and those who work closely with data. Learning data analysis and will help you bring your data skills to the next level.

This course starts by walking you through creating your own data and bringing data into Excel from various sources. You’ll learn the basics of SQL syntax and how to connect it to a Microsoft SQL Server Database using Excel’s data connection tools. You will discover how to spot bad data and strategies to clean that data to make it useful to you. Next, you'll learn to create custom columns, identify key metrics, and make decisions based on business rules. You’ll create macros using VBA and use Excel 2013’s shiny new macros. Finally, at the end of the course, you'll be provided with useful shortcuts and tips, enabling you to do efficient data analysis and business modeling with Excel 2013.

Working in a hands-on learning environment, led by our Excel Data Analysis expert instructor, students will learn about and explore:

* Create formulas to help you analyze and explain findings
* Develop interactive spreadsheets that will impress your audience and give them the ability to slice and dice data
* A step-by-step guide to learn various ways to model data for businesses with the help of Excel 2013

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Discover what Excel formulas are all about and how to use them in your spreadsheet development
* Identify bad data and learn cleaning strategies
* Create interactive spreadsheets that engage and appeal to your audience
* Leverage Excel’s powerful built-in tools to get the median, maximum, and minimum values of your data
* Build impressive tables and combine datasets using Excel’s built-in functionality
* Learn the powerful scripting language VBA, allowing you to implement your own custom solutions with ease

**Audience & Pre-Requisites**

This course is geared for attendees wants to learn various ways to model data for businesses with the help of Excel 2013

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Excel knowledge
* Good foundational mathematics or logic skills

**Course Agenda / Topics**

1. **Getting Data into Excel**

* Getting started with data
* Manually creating data
* Importing data from various sources

1. **Connecting to Databases**

* Reading a table from MSSQL – the Microsoft SQL Server database
* Reading multiple tables from MSSQL
* Reading from MSSQL using SQL

1. **How to Clean Texts, Numbers, and Dates**

* Leading/trailing/in-between spaces
* Capitalization
* Duplicates
* Text to Columns
* Combine data from multiple columns into one column
* Fixing similar words
* Text to dates
* Text to numbers

1. **Using Formulas to Prepare Your Data for Analysis**

* How to create formulas
* Combining strings and numbers
* Using built-in functions
* If/else/then statements
* Comparing columns

1. **Analyzing Your Data Using Descriptive Statistics and Charts**

* Gathering data
* Preparing the data for analysis
* Analyzing our data

1. **Link Your Data Using Data Models**

* Gathering data
* Preparing data
* Analyzing data
* Data models

1. **A Primer on Using the Excel Solver**

* Activating the Excel Solver
* Modeling our linear programming problem
* Using the Excel Solver

1. **Learning VBA – Excel's Scripting Language**

* What is VBA?
* What is a macro?
* Opening the VBA Editor
* Your very first "Hello World" VBA script

1. **How to Build and Style Your Charts**

* Quick analysis charts
* Charting options
* Additional design options

1. **Creating Interactive Spreadsheets Using Tables and Slicers**

* What are slicers?
* Final tweaks

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Build a Career in Data Science

**Course Snapshot**

* **Course:** Build a Career in Data Science
* **Duration:** 4 days
* **Skill-level**: Foundation-level Data Science skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared to know you that what are the keys to a data scientist’s long-term success.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Build a Career in Data** Science is your guide to landing your first data science job and developing into a valued senior employee. By following clear and simple instructions, you’ll learn to craft an amazing resume and ace your interviews. In this demanding, rapidly changing field, it can be challenging to keep projects on track, adapt to company needs, and manage tricky stakeholders. You’ll love the insights on how to handle expectations, deal with failures, and plan your career path in the stories from seasoned data scientists included in the course**.**

Working in a hands-on learning environment, led by Data Science expert instructor, students will learn about and explore:

* You’ll explore the insights on how to handle expectations, deal with failures, and plan your career path in the stories from seasoned data scientists included in the course.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Creating a portfolio of data science projects
* Assessing and negotiating an offer
* Leaving gracefully and moving up the ladder
* Interviews with professional data scientists

**Audience & Pre-Requisites**

This course is for readers who want to begin or advance a data science career.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **What is data science?**

* Different types of data science jobs
* Choosing your path
* Interview with Robert Chang, Data Scientist at Airbnb

1. **Data science companies**

* MTC – the Massive Tech Company
* Handbag LOVE – the established retailer
* Seg-Metra – the early-stage startup
* Videory – the late-stage, successful tech start-up
* Global Aerospace Dynamics (GAD) – the massive government contractor
* Putting it all together
* Interview with Randy Au, Quantitative User Experience Researcher at Google

1. **Getting the Skills**

* Earning a data science degree
* Going through a bootcamp
* Getting data science work within your company
* Teaching yourself
* Making the choice
* Interview with Julia Silge, Data Scientist at Stack Overflow

1. **Building a Portfolio**

* Creating a project
* Starting a Blog
* Example Projects
* Interview with David Robinson, Data Insights Engineering Manager at Flatiron Health

1. **The Search: Identifying the Right Job for You**

* Finding jobs
* Deciding which jobs to apply for
* Interview with Jesse Mostipak, Managing Director of Data Science at Teaching Trust

1. **The Application: Resumes and Cover Letters**

* Resume: the basics
* Cover letters: the basics
* Tailoring
* Referrals
* Interview with Kristen Kehrer, a data science instructor and course creator

1. **The interview: what to expect and how to handle it**

* What do companies want?
* The interview processes
* Step one: the initial phone screen interview
* Step two: the on-site interview
* Step three: the case study
* Step four: the final interview
* The offer
* Interview with Ryan Williams, Senior Decision Scientist at Starbucks

1. **The Offer: Knowing What to Accept**

* The process
* Receiving the offer
* Negotiation
* Negotiation Tactics
* How to pick between two “good” job offers
* Interview with Brooke Watson Madubuonwu, a Senior Data Scientist at the ACLU

1. **The First Months on the Job**

* The First Month
* Becoming productive
* If you’re the first data scientist
* When it’s not what was promised
* Interview with Jarvis Miller, Data Scientist at Spotify

1. **Making an effective analysis**

* The request
* The analysis plans
* Doing the analysis
* Wrapping it up
* Interview with Hilary Parker, a Data Scientist at Stitch Fix

1. **Deploying a model into production**

* What is deploying to production anyway?
* Making the production system
* Keeping the system running
* Wrapping up
* Interview with Heather Nolis, a Machine Learning Engineer at T-Mobile

1. **Working with stakeholders**

* Types of stakeholders
* Working with stakeholders
* Prioritizing work
* Concluding remarks
* Interview with Sade Snowden-Akintunde, a Data Scientist at Etsy

1. **When your data science project fails**

* Why data science projects fail
* Managing risk
* What you can do when your projects failed
* Interview with Michelle Keim, Head of Data Science & Machine Learning at Pluralsight

1. **Joining the Data Science Community**

* Growing your portfolio
* Attending Conferences
* Giving talks
* Contributing to open source
* Recognizing and avoiding burnout
* Interview with Renee Teate, Director of Data Science at HelioCampus

1. **Leaving your job gracefully**

* Deciding to leave
* How the job search differs after your first job
* Giving notice
* Interview with Amanda Casari, Engineering Manager at Google

1. **Moving up the ladder**

* The management tracks
* Principal data scientist track
* Switching to independent consulting
* Choosing your path
* Interview with Angela Bassa, Head of Data Science, Data Engineering, and Machine Learning at iRobot

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Science Boot camp

**Course Snapshot**

* **Course:** Data Science Boot camp
* **Duration:** 6 days
* **Skill-level**: Foundation-level Data Science Boot camp skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared to test and build your knowledge of Python and learn to handle the kind of open-ended problems that professional data scientists work on daily. Downloadable data sets and thoroughly-explained solutions help you lock in what you’ve learned, building your confidence and making you ready for an exciting new data science career.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Data Science Boot camp** is a comprehensive set of challenging projects carefully designed to grow your data science skills from novice to master. Veteran data scientist Leonard Apeltsin sets 10 increasingly difficult exercises that test your abilities against the kind of problems you’d encounter in the real-world. As you solve each challenge, you’ll acquire and expand the data science and Python skills you’ll use as a professional data scientist. Ranging from text processing to machine learning, each project comes complete with a unique downloadable data set and a fully-explained step-by-step solution. Because these projects come from Dr. Apelstin’s vast experience, each solution highlights the most likely failure points along with practical advice for getting past unexpected pitfalls. When you wrap up these 10 awesome exercises, you’ll have a diverse relevant skill set that’s transferable to working in industry.

Working in a hands-on learning environment, led by Data Science Boot camp expert instructor, students will learn about and explore:

* Visualize complex multi-variable datasets
* Train a decision tree machine learning algorithm

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* 10 in-depth Python exercises with full downloadable data sets
* Web scraping for text and images
* Organize data sets with clustering algorithms
* Visualize complex multi-variable datasets
* Train a decision tree machine learning algorithm

**Audience & Pre-Requisites**

This course is for readers who know the basics of Python. No prior data science or machine learning skills required.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Computing Probabilities Using Python**

* Sample Space Analysis: An Equation-Free Approach for Measuring Uncertainty in Outcomes
* Computing Non-Trivial Probabilities
* Computing Probabilities Over Interval Ranges

1. **Plotting Probabilities Using Matplotlib**

* Basic Matplotlib Plots
* Plotting Coin-Flip Probabilities

1. **Running Random Simulations in NumPy**

* Simulating Random Coin-Flips and Dice-Rolls Using NumPy
* Computing Confidence Intervals Using Histograms and NumPy Arrays
* Leveraging Confidence Intervals to Analyze a Biased Deck of Cards
* Using Permutations to Shuffle Cards

1. **Case Study 1 Solution**

* Overview
* Predicting Red Cards within a Shuffled Deck
* Optimizing Strategies using the Sample Space for a 10-Card Deck
* Key Takeaways
* Part 2. Case Study 2: Assessing Online Ad-Clicks for Significance

1. **5 Basic Probability and Statistical Analysis Using SciPy**

* Exploring the Relationships between Data and Probability Using SciPy
* Mean as a Measure of Centrality
* Variance as a Measure of Dispersion

1. **Making Predictions Using the Central Limit Theorem and SciPy**

* Manipulating the Normal Distribution Using SciPy
* 6.2 Determining Mean and Variance of a Population through Random Sampling
* 6.3 Making Predictions Using Mean

1. **Statistical Hypothesis Testing**

* Assessing the Divergence Between Sample Mean and Population Mean
* Data Dredging: Coming to False Conclusions through Oversampling
* Bootstrapping with Replacement: Testing a Hypothesis When the Population Variance is Unknown
* Permutation Testing: Comparing Means of Samples when the Population Parameters are Unknown

1. **Analyzing Tables Using Pandas**

* Storing Tables Using Basic Python
* Exploring Tables Using Pandas
* Retrieving Table Columns
* Retrieving Table Rows
* Modifying Table Rows and Columns
* Saving and Loading Table Data
* Visualizing Tables Using Seaborn

1. **Case Study 2 Solution Overview**

* Processing the Ad-Click Table in Pandas
* Computing P-values from Differences in Means
* Determining Statistical Significance
* Shades of Blue: A Real-Life Cautionary Tale
* Key Takeaways
* Part 3. Case Study 3: Tracking Disease Outbreaks Using News Headlines

1. **Clustering Data into Groups**

* Using Centrality to Discover Clusters
* K-Means: A Clustering Algorithm for Grouping Data into K Central Groups
* Using the Elbow Method
* Using Density to Discover Clusters
* DBSCAN: A Clustering Algorithm for Grouping Data Based on Spatial Density
* Analyzing Clusters Using Pandas

1. **Geographic Location Visualization and Analysis**

* The Great-Circle Distance: A Metric for Computing Distances Between 2 Global Points
* Plotting Maps Using Base map
* Location Tracking Using GeoNamesCache
* Matching Location Names in Text

1. **Case Study 3 Solution**

* Overview
* Extracting Locations from Headline Data
* Visualizing and Clustering the Extracted Location Data
* Extracting Insights from Location Clusters
* Key Takeaways
* Part 4. Case Study 4: Using Online Job Postings to Improve Your Data Science Resume

1. **Measuring Text Similarities**

* Simple Text Comparison
* Vectorizing Texts Using Word Counts
* Matrix Multiplication for Efficient Similarity Calculation
* Computational Limits of Matrix Multiplication

1. **Dimension Reduction of Matrix Data**

* Clustering 2D Data in 1-Dimension
* Dimension Reduction Using PCA and Scikit-Learn
* Clustering 4D Data in 2-Dimensions
* Computing Principal Components Without Rotation
* Efficient Dimension Reduction Using SVD and Scikit-Learn

1. **NLP Analysis of Large Text Datasets**
2. **Extracting Text from Web Pages**

* The Structure of HTML Documents
* Parsing HTML using Beautiful Soup
* Downloading and Parsing Online Data

1. **Case Study 4 Solution**

* Overview
* Extracting Skill Requirements from Job Posting Data
* Filtering Jobs by Relevance
* Clustering Skills in Relevant Job Postings
* Conclusion
* Key Takeaways

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Healthcare Analytics Made Simple

Course Snapshot

* **Course:** Healthcare Analytics Made Simple
* **Duration:** 2 days
* **Skill-level**: Foundation-level Healthcare Analytics Made Simple for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to add a touch of data analytics to your healthcare systems and get insightful outcomes
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In recent years, machine learning technologies and analytics have been widely utilized across the healthcare sector. Healthcare Analytics Made Simple bridges the gap between practicing doctors and data scientists. It equips the data scientists’ work with healthcare data and allows them to gain better insight from this data in order to improve healthcare outcomes.

This course is a complete overview of machine learning for healthcare analytics, briefly describing the current healthcare landscape, machine learning algorithms, and Python and SQL programming languages. The step-by-step instructions teach you how to obtain real healthcare data and perform descriptive, predictive, and prescriptive analytics using popular Python packages such as pandas and scikit-learn. The latest research results in disease detection and healthcare image analysis are reviewed.

.

Working in a hands-on learning environment, led by our Healthcare expert instructor, students will learn about and explore:

* Perform healthcare analytics with Python and SQL
* Build predictive models on real healthcare data with pandas and scikit-learn
* Use analytics to improve healthcare performance

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Gain valuable insight into healthcare incentives, finances, and legislation
* Discover the connection between machine learning and healthcare processes
* Use SQL and Python to analyze data
* Measure healthcare quality and provider performance
* Identify features and attributes to build successful healthcare models
* Build predictive models using real-world healthcare data
* Become an expert in predictive modeling with structured clinical data
* See what lies ahead for healthcare analytics

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to know how to use Python for healthcare data analysis, how to import, collect, clean, and refine data from electronic health record (EHR) surveys, and how to make predictive models with this data through real-world algorithms and code examples

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Introduction to Healthcare Analytics**

* What is healthcare analytics?
* Foundations of healthcare analytics
* History of healthcare analytics
* Examples of healthcare analytics
* Exploring the software

1. **Healthcare Foundations**

* Healthcare delivery in the US
* Patient data – the journey from patient to computer
* Standardized clinical codesets
* Breaking down healthcare analytics

1. **Machine Learning Foundations**

* Model frameworks for medical decision making
* Machine learning pipeline

1. **Computing Foundations – Databases**

* Introduction to databases
* Data engineering with SQL – an example case
* Case details – predicting mortality for a cardiology practice
* Starting an SQLite session
* Data engineering, one table at a time with SQL

1. **Computing Foundations – Introduction to Python**

* Variables and types
* Data structures and containers
* Programming in Python – an illustrative example
* Introduction to pandas
* Introduction to scikit-learn
* Additional analytics libraries

1. **Measuring Healthcare Quality**

* Introduction to healthcare measures
* US Medicare value-based programs
* The Hospital Value-Based Purchasing (HVBP) program
* The Hospital Readmission Reduction (HRR) program
* The Hospital-Acquired Conditions (HAC) program
* The End-Stage Renal Disease (ESRD) quality incentive program
* The Skilled Nursing Facility Value-Based Program (SNFVBP)
* The Home Health Value-Based Program (HHVBP)
* The Merit-Based Incentive Payment System (MIPS)
* Other value-based programs
* Comparing dialysis facilities using Python
* Comparing hospitals

1. **Making Predictive Models in Healthcare**

* Introduction to predictive analytics in healthcare
* Our modeling task – predicting discharge statuses for ED patients
* Obtaining the dataset
* Starting a Jupyter session
* Importing the dataset
* Making the response variable
* Splitting the data into train and test sets
* Preprocessing the predictor variables
* Final preprocessing steps
* Building the models
* Using the models to make predictions
* Improving our models

1. **Healthcare Predictive Models – A Review**

* Predictive healthcare analytics – state of the art
* Overall cardiovascular risk
* Congestive heart failure
* Cancer
* Readmission prediction
* Other conditions and events

1. **The Future – Healthcare and Emerging Technologies**

* Healthcare analytics and the internet
* Healthcare and deep learning
* Obstacles, ethical issues, and limitations

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning for Healthcare Analytics Projects

**Course Snapshot**

* **Course:** Machine Learning for Healthcare Analytics Projects
* **Duration:** 1 days
* **Skill-level**: Foundation-level machine learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to create real-world machine learning solutions using NumPy, pandas, matplotlib, and scikit-learn
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Machine Learning (ML)** has changed the way organizations and individuals use data to improve the efficiency of a system. ML algorithms allow strategists to deal with a variety of structured, unstructured, and semi-structured data. Machine Learning for Healthcare Analytics Projects is packed with new approaches and methodologies for creating powerful solutions for healthcare analytics. This course will teach you how to implement key machine learning algorithms and walk you through their use cases by employing a range of libraries from the Python ecosystem. You will build five end-to-end projects to evaluate the efficiency of Artificial Intelligence (AI) applications for carrying out simple-to-complex healthcare analytics tasks. With each project, you will gain new insights, which will then help you handle healthcare data efficiently. As you make your way through the course, you will use ML to detect cancer in a set of patients using support vector machines (SVMs) and k-Nearest neighbors (KNN) models. In the final lessons, you will create a deep neural network in Keras to predict the onset of diabetes in a huge dataset of patients. You will also learn how to predict heart diseases using neural networks.

Working in a hands-on learning environment, led by our Python expert instructor, students will learn about and explore:

* Develop a range of healthcare analytics projects using real-world datasets
* Implement key machine learning algorithms using a range of libraries from the Python ecosystem
* Accomplish intermediate-to-complex tasks by building smart AI applications using neural network methodologies

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Explore super imaging and natural language processing (NLP) to classify DNA sequencing
* Detect cancer based on the cell information provided to the SVM
* Apply supervised learning techniques to diagnose autism spectrum disorder (ASD)
* Implement a deep learning grid and deep neural networks for detecting diabetes
* Analyze data from blood pressure, heart rate, and cholesterol level tests using neural networks
* Use ML algorithms to detect autistic disorders

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to know how to address long-standing challenges, provide specialized solutions for how to deal with them, and carry out a range of cognitive tasks in the healthcare domain..

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Breast Cancer Detection**

* Objective of this project
* Detecting breast cancer with SVM and KNN models
* Training models

1. **Diabetes Onset Detection**

* Detecting diabetes using a grid search
* Introduction to the dataset
* Building our Keras model
* Performing a grid search using scikit-learn
* Reducing overfitting using dropout regularization
* Finding the optimal hyperparameters
* Optimizing the number of neurons
* Generating predictions using optimal hyperparameters

1. **DNA Classification**

* Classifying DNA sequences

1. **Diagnosing Coronary Artery Disease**

* The dataset
* Fixing missing data
* Splitting the dataset
* Training the neural network
* A comparison of categorical and binary problems

1. **Autism Screening with Machine Learning**

* ASD screening using machine learning
* Introducing the dataset
* Splitting the dataset into training and testing datasets
* Building the network
* Testing the network

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning for Finance

**Course Snapshot**

* **Course:** Machine Learning for Finance
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning for Finance skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to get a guide to advances in machine learning for financial professionals, with working Python code
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Machine Learning for Finance** explores new advances in machine learning and shows how they can be applied across the financial sector, including insurance, transactions, and lending. This course explains the concepts and algorithms behind the main machine learning techniques and provides example Python code for implementing the models yourself. The course is based on Jannes Klaas’ experience of running machine learning training courses for financial professionals. Rather than providing ready-made financial algorithms, the course focuses on advanced machine learning concepts and ideas that can be applied in a wide variety of ways. The course systematically explains how machine learning works on structured data, text, images, and time series. You'll cover generative adversarial learning, reinforcement learning, debugging, and launching machine learning products. Later lessons will discuss how to fight bias in machine learning. The course ends with an exploration of Bayesian inference and probabilistic programming.

Working in a hands-on learning environment, led by our Machine Learning for Finance expert instructor, students will learn about and explore:

* Explore advances in machine learning and how to put them to work in financial industries
* Gain expert insights into how machine learning works, with an emphasis on financial applications
* Discover advanced machine learning approaches, including neural networks, GANs, and reinforcement learning

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Apply machine learning to structured data, natural language, photographs, and written text
* Understand how machine learning can help you detect fraud, forecast financial trends, analyze customer sentiments, and more
* Implement heuristic baselines, time series, generative models, and reinforcement learning in Python, scikit-learn, Keras, and TensorFlow
* Delve into neural networks, and examine the uses of GANs and reinforcement learning
* Debug machine learning applications and prepare them for launch
* Address bias and privacy concerns in machine learning

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to get a guide to advances in machine learning for financial professionals, with working Python code

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Neural Networks and Gradient-Based Optimization**

* Neural Networks and Gradient-Based Optimization
* Our journey in this course
* What is machine learning?
* Supervised learning
* Unsupervised learning
* Reinforcement learning
* Setting up your workspace
* Using Kaggle kernels
* Using the AWS deep learning AMI
* Approximating functions
* A forward pass
* A logistic regressor
* Optimizing model parameters
* Measuring model loss
* A deeper network
* A brief introduction to Keras
* Tensors and the computational graph

1. **Applying Machine Learning to Structured Data**

* Applying Machine Learning to Structured Data
* The data
* Heuristic, feature-based, and E2E models
* The machine learning software stack
* The heuristic approach
* The feature engineering approach
* Preparing the data for the Keras library
* Creating predictive models with Keras
* A brief primer on tree-based methods
* E2E modeling

1. **Utilizing Computer Vision**

* Utilizing Computer Vision
* Convolutional Neural Networks
* Filters on color images
* The building blocks of ConvNets in Keras
* More bells and whistles for our neural network
* Working with big image datasets
* Working with pretrained models
* The modularity tradeoff
* Computer vision beyond classification

1. **Understanding Time Series**

* Understanding Time Series
* Visualization and preparation in pandas
* Fast Fourier transformations
* Autocorrelation
* Establishing a training and testing regime
* A note on backtesting
* Median forecasting
* ARIMA
* Kalman filters
* Forecasting with neural networks
* Conv1D
* Dilated and causal convolution
* Simple RNN
* LSTM
* Recurrent dropout
* Bayesian deep learning

1. **Parsing Textual Data with Natural Language Processing**

* Parsing Textual Data with Natural Language Processing
* An introductory guide to spaCy
* Named entity recognition
* Part-of-speech (POS) tagging
* Rule-based matching
* Regular expressions
* A text classification task
* Preparing the data
* Bag-of-words
* Topic modeling
* Word embeddings
* Document similarity with word embeddings
* A quick tour of the Keras functional API
* Attention
* Seq2seq models

1. **Using Generative Models**

* Using Generative Models
* Understanding autoencoders
* Visualizing latent spaces with t-SNE
* Variational autoencoders
* VAEs for time series
* GANs
* Using less data – active learning
* SGANs for fraud detection

1. **Reinforcement Learning for Financial Markets**

* Reinforcement Learning for Financial Markets
* Catch – a quick guide to reinforcement learning
* Markov processes and the bellman equation – A more formal introduction to RL
* Advantage actor-critic models
* Evolutionary strategies and genetic algorithms
* Practical tips for RL engineering
* Frontiers of RL

1. **Privacy, Debugging, and Launching Your Products**

* Privacy, Debugging, and Launching Your Products
* Debugging data
* Debugging your model
* Deployment

1. **Fighting Bias**

* Fighting Bias
* Sources of unfairness in machine learning
* Legal perspectives
* Observational fairness
* Training to be fair
* Causal learning
* Interpreting models to ensure fairness
* Unfairness as complex system failure
* A checklist for developing fair models

1. **Bayesian Inference and Probabilistic Programming**

* Bayesian Inference and Probabilistic Programming
* An intuitive guide to Bayesian inference

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Go Machine Learning Projects

**Course Snapshot**

* **Course:** Go Machine Learning Projects
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Work through exciting projects to explore the capabilities of Go and Machine Learning
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Go is the perfect language for machine learning;** it helps to clearly describe complex algorithms, and also helps developers to understand how to run efficient optimized code. This course will teach you how to implement machine learning in Go to make programs that are easy to deploy and code that is not only easy to understand and debug, but also to have its performance measured. The course begins by guiding you through setting up your machine learning environment with Go libraries and capabilities. You will then plunge into regression analysis of a real-life house pricing dataset and build a classification model in Go to classify emails as spam or ham. Using Gonum, Gorgonia, and STL, you will explore time series analysis along with decomposition and clean up your personal Twitter timeline by clustering tweets. In addition to this, you will learn how to recognize handwriting using neural networks and convolutional neural networks. Lastly, you'll learn how to choose the most appropriate machine learning algorithms to use for your projects with the help of a facial detection project. By the end of this course, you will have developed a solid machine learning mindset, a strong hold on the powerful Go toolkit, and a sound understanding of the practical implementations of machine learning algorithms in real-world projects.

Working in a hands-on learning environment, led by our Go Machine Learning expert instructor, students will learn about and explore:

* Explore ML tasks and Go’s machine learning ecosystem
* Implement clustering, regression, classification, and neural networks with Go
* Get to grips with libraries such as Gorgonia, Gonum, and GoCv for training models in Go

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Set up a machine learning environment with Go libraries
* Use Gonum to perform regression and classification
* Explore time series models and decompose trends with Go libraries
* Clean up your Twitter timeline by clustering tweets
* Learn to use external services for your machine learning needs
* Recognize handwriting using neural networks and CNN with Gorgonia
* Implement facial recognition using GoCV and OpenCV

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Work through exciting projects to explore the capabilities of Go and Machine Learning

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **How to Solve All Machine Learning Problems**

* How to Solve All Machine Learning Problems
* What is a problem?
* What is an algorithm?
* What is machine learning?
* Do you need machine learning?
* The general problem solving process
* What is a model?
* On writing and lesson organization
* Why Go?
* Quick start
* Functions
* Variables

1. **Linear Regression - House Price Prediction**

* Linear Regression - House Price Prediction
* The project
* Exploratory data analysis
* Linear regression
* Discussion and further work

1. **Classification - Spam Email Detection**

* Classification - Spam Email Detection
* The project
* Exploratory data analysis
* The classifier
* Naive Bayes
* Implementating the classifier
* Putting it all together

1. **Decomposing CO2 Trends Using Time Series Analysis**

* Decomposing CO2 Trends Using Time Series Analysis
* Exploratory data analysis
* Decomposition
* Forecasting

1. **Clean Up Your Personal Twitter Timeline by Clustering Tweets**

* Clean Up Your Personal Twitter Timeline by Clustering Tweets
* The project
* K-means
* DBSCAN
* Data acquisition
* Exploratory data analysis
* Data massage
* Clustering
* Real data
* The program
* Tweaking the program

1. **Neural Networks - MNIST Handwriting Recognition**

* Neural Networks - MNIST Handwriting Recognition
* A neural network
* Linear algebra 101
* Learning
* The project
* Training the neural network
* Cross-validation

1. **Convolutional Neural Networks - MNIST Handwriting Recognition**

* Convolutional Neural Networks - MNIST Handwriting Recognition
* Everything you know about neurons is wrong
* Neural networks – a redux
* The project
* CNNs
* Describing a CNN
* Running the neural network
* Testing

1. **Basic Facial Detection**

* Basic Facial Detection
* What is a face?
* PICO
* GoCV
* Pigo
* Face detection program
* Evaluating algorithms

1. **Hot Dog or Not Hot Dog - Using External Services**

* Hot Dog or Not Hot Dog - Using External Services
* MachineBox
* What is MachineBox?
* The project
* The results
* What does this all mean?
* Why MachineBox?

1. **What's Next?**

* What's Next?
* What should the reader focus on?
* The researcher, the practitioner, and their stakeholder
* What did this course not cover?
* Where can I learn more?

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning Projects for Mobile Applications

**Course Snapshot**

* **Course:** Machine Learning Projects for Mobile Applications
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Build Android and iOS applications using TensorFlow Lite and Core ML
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning is a technique that focuses on developing computer programs that can be modified when exposed to new data. We can make use of it for our mobile applications and this course will show you how to do so. The course starts with the basics of machine learning concepts for mobile applications and how to get well equipped for further tasks. You will start by developing an app to classify age and gender using Core ML and Tensorflow Lite. You will explore neural style transfer and get familiar with how deep CNNs work. We will also take a closer look at Google’s ML Kit for the Firebase SDK for mobile applications. You will learn how to detect handwritten text on mobile. You will also learn how to create your own Snapchat filter by making use of facial attributes and OpenCV. You will learn how to train your own food classification model on your mobile; all of this will be done with the help of deep learning techniques. Lastly, you will build an image classifier on your mobile, compare its performance, and analyze the results on both mobile and cloud using TensorFlow Lite with an RCNN. By the end of this course, you will not only have mastered the concepts of machine learning but also learned how to resolve problems faced while building powerful apps on mobiles using TensorFlow Lite, Caffe2, and Core ML.

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Explore machine learning using classification, analytics, and detection tasks.
* Work with image, text and video datasets to delve into real-world tasks
* Build apps for Android and iOS using Caffe, Core ML and Tensorflow Lite

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Demystify the machine learning landscape on mobile
* Age and gender detection using TensorFlow Lite and Core ML
* Use ML Kit for Firebase for in-text detection, face detection, and barcode scanning
* Create a digit classifier using adversarial learning
* Build a cross-platform application with face filters using OpenCV
* Classify food using deep CNNs and TensorFlow Lite on iOS

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Build Android and iOS applications using TensorFlow Lite and Core ML

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Mobile Landscapes in Machine Learning**

* Mobile Landscapes in Machine Learning
* Machine learning basics
* TensorFlow Lite and Core ML
* TensorFlow Lite
* Core ML

1. **CNN Based Age and Gender Identification Using Core ML**

* CNN Based Age and Gender Identification Using Core ML
* Age, gender, and emotion prediction
* Convolutional Neural Networks
* The implementation on iOS using Core ML

1. **Applying Neural Style Transfer on Photos**

* Applying Neural Style Transfer on Photos
* Artistic neural style transfer
* Building the applications

1. **Deep Diving into the ML Kit with Firebase**

* Deep Diving into the ML Kit with Firebase
* ML Kit basics
* Face detection
* Barcode scanner
* Text recognition

1. **A Snapchat-Like AR Filter on Android**

* A Snapchat-Like AR Filter on Android
* MobileNet models
* Building the Android application

1. **Handwritten Digit Classifier Using Adversarial Learning**

* Handwritten Digit Classifier Using Adversarial Learning
* Generative Adversarial Networks
* Understanding the MNIST database
* Building the TensorFlow model
* Training the neural network

1. **Face-Swapping with Your Friends Using OpenCV**

* Face-Swapping with Your Friends Using OpenCV
* Understanding face-swapping

1. **Classifying Food Using Transfer Learning**

* Classifying Food Using Transfer Learning
* Transfer learning
* Training our own TensorFlow model

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# IoT Solutions with Blockchain

**Course Snapshot**

* **Course:** IoT Solutions with Blockchain
* **Duration:** 2 days
* **Skill-level**: Foundation-level On IoT Solutions with Blockchain skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Integrate an end-to-end logistic chain using IBM Blockchain and IoT platforms
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Blockchain** has been the hot topic of late thanks to cryptocurrencies. To make matters more interesting, the financial market is looking for ways to reduce operational costs and generate new business models, and this is where blockchain solutions come into the picture. In addition to this, with Internet of Things (IoT) trending and Arduino, Raspberry Pi, and other devices flooding the market, you can now create cheap devices even at home. Hands-On IoT Solutions with Blockchain starts with an overview of IoT concepts in the current business scenario. It then helps you develop your own device on the IBM Watson IoT platform and create your first IoT solution using Watson and Intel Edison. Once you are familiar with IoT, you will learn about Blockchain technology and its use cases. You will also work with the Hyperledger framework and develop your own Blockchain network. As you progress through the lesson, you'll work with problem statements and learn how to design your solution architecture so that you can create your own integrated Blockchain and IoT solution.

Working in a hands-on learning environment, led by our IoT Solutions with Blockchain expert instructor, students will learn about and explore:

* Explore practical implementation of ledger technology in the IoT architecture
* Study security best practices for your smart devices
* Understand Blockchain implementation for end-to-end IoT solutions

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the key roles of IoT in the current market
* Study the different aspects of IBM Watson IoT platform
* Create devices, gateways, and applications connected to the platform
* Explore the fundamentals of Blockchain
* Define good use cases for Blockchain
* Discover the Hyperledger Fabric and Composer frameworks
* Develop an IBM Watson IoT application using a Intel Edison
* Integrate IoT with the Blockchain platform

**Audience & Pre-Requisites**

This course is geared for attendees who wants to know that how to implement end-to-end Blockchain solutions with IoT using the IBM Cloud platform and mastered the convergence of IoT and Blockchain technology and exploited the best practices and drivers to develop a bulletproof integrated solution.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, and Blockchain knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Understanding IoT and Developing Devices on the IBM Watson IoT Platform**

* What is IoT?
* Common business use cases of IoT
* Technical elements in IoT
* IBM Watson IoT Platform features
* Creating your first IoT solution

1. **Creating Your First IoT Solution**

* Technical requirements
* The first IoT solution – the gardening solution
* Coding the device firmware
* Creating the backend application

1. **Explaining Blockchain Technology and Working with Hyperledger**

* What is blockchain?
* Blockchain and Hyperledger
* Selecting a good use case

1. **Creating Your Own Blockchain Network**

* Prerequisites
* Creating your own blockchain network with Hyperledger Composer
* Installing your own blockchain network using Hyperledger Fabric and Composer

1. **Addressing Food Safety - Building around the Blockchain**

* Regulations, challenges, and concerns in the modern food chain
* How blockchain and IoT can help in a food chain

1. **Designing the Solution Architecture**

* The business of food
* The technological approach
* Software components

1. **Creating Your Blockchain and IoT Solution**

* Technical requirements
* Solution overview
* Creating a blockchain network
* Creating the IoT part of the solution
* End-to-end testing

1. **The IoT, Blockchain, and Industry 4.0**

* Industry 4.0
* Cloud computing as an innovation platform
* The IoT
* Blockchain – simplifying business chains

1. **Best Practices for Developing Blockchain and IoT Solutions**

* Developing cloud applications
* Blockchain development using Hyperledger Composer

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Music Generation with Magenta

**Course Snapshot**

* **Course:** Music Generation with Magenta
* **Duration:** 3 days
* **Skill-level**: Foundation-level Magenta skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants design and use machine learning models for music generation using Magenta and make them interact with existing music creation tools
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The importance of **machine learning (ML)** in art is growing at a rapid pace due to recent advancements in the field, and Magenta is at the forefront of this innovation. With this course, you’ll follow a hands-on approach to using ML models for music generation, learning how to integrate them into an existing music production workflow. Complete with practical examples and explanations of the theoretical background required to understand the underlying technologies, this course is the perfect starting point to begin exploring music generation. The course will help you learn how to use the models in Magenta for generating percussion sequences, monophonic and polyphonic melodies in MIDI, and instrument sounds in raw audio. Through practical examples and in-depth explanations, you’ll understand ML models such as RNNs, VAEs, and GANs. Using this knowledge, you’ll create and train your own models for advanced music generation use cases, along with preparing new datasets. Finally, you’ll get to grips with integrating Magenta with other technologies, such as digital audio workstations (DAWs), and using Magenta.js to distribute music generation apps in the browser..

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Learn how machine learning, deep learning, and reinforcement learning are used in music generation
* Generate new content by manipulating the source data using Magenta utilities, and train machine learning models with it
* Explore various Magenta projects such as Magenta Studio, MusicVAE, and NSynth

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use RNN models in Magenta to generate MIDI percussion, and monophonic and polyphonic sequences
* Use WaveNet and GAN models to generate instrument notes in the form of raw audio
* Employ Variational Autoencoder models like MusicVAE and GrooVAE to sample, interpolate, and humanize existing sequences
* Prepare and create your dataset on specific styles and instruments
* Train your network on your personal datasets and fix problems when training networks
* Apply MIDI to synchronize Magenta with existing music production tools like DAWs

**Audience & Pre-Requisites**

This course is geared for attendees with Machine Learning on Google Cloud Platform skills who wish to l be well-versed with Magenta and have developed the skills you need to use ML models for music generation in your own style.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Section 1: Introduction to Artwork Generation**
2. **Introduction to Magenta and Generative Art**

* Technical requirements
* Overview of generative art
* New techniques with machine learning
* Google's Magenta and TensorFlow in music generation
* Installing Magenta and Magenta for GPU
* Installing the music software and synthesizers
* Installing the code editing software
* Generating a basic MIDI file

1. **Section 2: Music Generation with Machine Learning**
2. **Generating Drum Sequences with the Drums RNN**

* Technical requirements
* The significance of RNNs in music generation
* Using the Drums RNN on the command line
* Using the Drums RNN in Python

1. **Generating Polyphonic Melodies**

* Technical requirements
* LSTM for long-term dependencies
* Generating melodies with the Melody RNN
* Generating polyphony with the Polyphony RNN and Performance RNN

1. **Latent Space Interpolation with MusicVAE**

* Technical requirements
* Continuous latent space in VAEs
* Score transformation with MusicVAE and GrooVAE
* Understanding TensorFlow code

1. **Audio Generation with NSynth and GANSynth**

* Technical requirements
* Learning about WaveNet and temporal structures for music
* Neural audio synthesis with NSynth
* Using GANSynth as a generative instrument

1. **Section 3: Training, Learning, and Generating a Specific Style**
2. **Data Preparation for Training**

* Technical requirements
* Looking at existing datasets
* Building a dance music dataset
* Building a jazz dataset
* Preparing the data using pipelines

1. **Training Magenta Models**

* Technical requirements
* Choosing the model and configuration
* Training and tuning a model
* Using Google Cloud Platform

1. **Section 4: Making Your Models Interact with Other Applications**
2. **Magenta in the Browser with Magenta.js**

* Technical requirements
* Introducing Magenta.js and TensorFlow.js
* Creating a Magenta.js web application
* Making Magenta.js interact with other apps

1. **Making Magenta Interact with Music Applications**

* Technical requirements
* Sending MIDI to a DAW or synthesizer
* Looping the generated MIDI
* Using Magenta as a standalone application with Magenta Studio

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Practical Recommender Systems

Course Snapshot

* **Course:** Practical Recommender Systems
* **Duration:** 4 days
* **Skill-level**: Foundation- Practical Recommender Systems for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending be able to read code in a programming language such as Python or Java, you should understand an SQL query, and you should have a basic understanding of higher math and statistics.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** Delivery Format: This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Are you envious when Amazon recommends its products or when Netflix is spot-on with a recommendation for a user? Then here’s your chance to learn how to add these skills to your repertoire. Reading this course will give you an understanding of what recommender systems are and how to apply them in practice. To make a recommender work, many things need to perform in concert. You need to understand how to collect data from your users and how to interpret it, and you need a toolbox of different recommender algorithms so you can choose the best one for your particular scenario. Most importantly, you need to understand how to evaluate whether your recommender system is doing its job well. All this and more is hidden within this course.

Working in a hands-on learning environment, led by our Recommendation Systems expert instructor, students will learn about and explore:

* you should be able to read code in a programming language such as Python or Java
* you should understand an SQL query, and you should have a basic understanding of higher math and statistics.
* Figures and code listings that explain concepts can get you only so far.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* How to understand users and their behavior, and covers ways to collect data from users.
* Introduces web analytics and shows how you can implement a dashboard where you can keep track of your recommenders.
* How behavioral data can be transformed into ratings.
* Outlines the problem of new users and products and gives simple solutions.
* Discusses formulas for calculating similarity between users or content items such as movies.
* presents a way to mix types of recommenders.
* Introduces ranking algorithms and methods for learning to rank recommendations.
* looks at non-personalized recommendations.

Audience & Pre-Requisites

This course is geared for attendees with Recommendation Systems with Python skills and infrastructure needed to get them up and running. Managers will find this course useful to get an overview of what a recommender system is and how it can be used in practice.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

**1. What is a recommender?**

* Real-life recommendations
* Taxonomy of recommender systems
* Machine learning and the Netflix Prize
* The MovieGEEKs website
* Building a recommender system

**2. User behavior and how to collect it**

* How (I think) Netflix gathers evidence while you browse
* Finding useful user behavior
* Identifying users
* Getting visitor data from other sources
* The collector
* What users in the system are and how to model them

**3. Monitoring the system**

* Why adding a dashboard is a good idea
* Doing the analytics
* Personas
* MovieGEEKs dashboard

**4. Ratings and how to calculate them**

* User-item preferences
* Explicit or implicit ratings
* Revisiting explicit ratings
* What are implicit ratings?
* Calculating implicit ratings
* How to implement implicit ratings
* Less frequent items provide more value

**5. Non-personalized recommendations**

* What’s a non-personalized recommendation?
* How to make recommendations when you have no data
* Implementing the chart and the groundwork for the recommender system component
* Seeded recommendations

**6. The user (and content) who came in from the cold**

* What’s a cold start?
* Keeping track of visitors
* Addressing cold-start problems with algorithms
* Those who doesn’t ask, won’t know
* Using association rules to start recommending things fast

**7. Finding similarities among users and among content**

* Why similarity?
* Essential similarity functions
* k-means clustering
* Implementing similarities

**8. Collaborative filtering in the neighborhood**

* Collaborative filtering: A history lesson
* Calculating recommendations
* Calculating similarities
* Amazon’s algorithm to precalculate item similarity
* Ways to select the neighborhood
* Finding the right neighborhood
* Ways to calculate predicted ratings
* Prediction with item-based filtering
* Cold-start problems
* A few words on machine learning terms
* Collaborative filtering on the MovieGEEKs site
* What’s the difference between association rule recs and collaborative recs?
* Levers to fiddle with for collaborative filtering
* Pros and cons of collaborative filtering

**9. Evaluating and testing your recommender**

* Business wants lift, cross-sales, up-sales, and conversions
* Why is it important to evaluate?
* How to interpret user behavior
* What to measure
* Before implementing the recommender...
* Types of evaluation
* Offline evaluation
* Offline experiments
* Implementing the experiment in MovieGEEKs
* Evaluating the test set
* Online evaluation
* Continuous testing with exploit/explore

**10. Content-based filtering**

* Descriptive example
* Content-based filtering
* Content analyzer
* Extracting metadata from descriptions
* Finding important words with TF-IDF
* Topic modeling using the LDA
* Finding similar content
* Creating the user profile
* Content-based recommendations in MovieGEEKs
* Evaluation of the content-based recommender
* Pros and cons of content-based filtering

**11. Finding hidden genres with matrix factorization**

* Sometimes it’s good to reduce the amount of data
* Example of what you want to solve
* A whiff of linear algebra
* Matrix
* What’s factorization?
* Constructing the factorization using SVD
* Adding a new user by folding in
* How to do recommendations with SVD
* Baseline predictors
* Temporal dynamic
* Constructing the factorization using Funk SVD
* Root Mean Squared Error
* Gradient descent
* Stochastic gradient descent
* And finally, to the factorization
* Adding biases
* How to start and when to stop
* Doing recommendations with Funk SVD
* User vector
* The items the user likes
* Funk SVD implementation in MovieGEEKs
* What to do with outliers
* Keeping the model up to date
* Faster implementation
* Explicit vs. implicit data
* Evaluation
* Levers to fiddle with for Funk SVD

**12. Taking the best of all algorithms: Implementing hybrid recommenders**

* The confused world of hybrids
* The monolithic
* Mixing content-based features with behavioral data to improve collaborative filtering recommenders
* Mixed hybrid recommender
* The ensemble
* Switched ensemble recommender
* Weighted ensemble recommender
* Linear regression
* Feature-weighted linear stacking (FWLS)
* Meta features: Weights as functions
* The algorithm
* Implementation

**13. Ranking and learning to rank**

* Learning to rank an example at Foursquare
* Re-ranking
* What’s learning to rank again?
* The three types of LTR algorithms
* Bayesian Personalized Ranking
* Ranking with BPR
* Math magic (advanced wizardry)
* The BPR algorithm
* BPR with matrix factorization
* Implementation of BPR
* Doing the recommendations
* Evaluation
* Levers to fiddle with for BPR

**14. Future of recommender systems**

* Algorithms
* Context
* Human-computer interactions
* Choosing a good architecture
* What’s the future of recommender systems?
* User profiles
* context
* Algorithms
* Privacy
* Architecture
* Surprising recommendations

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Recommendation Systems with Python

Course Snapshot

* **Course:** Recommendation Systems with Python
* **Duration:** 2 days
* **Skill-level**: Foundation-level Recommendation Systems with Python for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to learn the tools and techniques required in building various kinds of powerful recommendation systems (collaborative, knowledge and content based) and deploying them to the web.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** Delivery Format: This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Recommendation systems** are at the heart of almost every internet business today; from Facebook to Netﬂix to Amazon. Providing good recommendations, whether it's friends, movies, or groceries, goes a long way in defining user experience and enticing your customers to use your platform. This course shows you how to do just that. You will learn about the different kinds of recommenders used in the industry and see how to build them from scratch using Python. No need to wade through tons of machine learning theory—you'll get started with building and learning about recommenders as quickly as possible. In this course, you will build an IMDB Top 250 clone, a content-based engine that works on movie metadata. You'll use collaborative filters to make use of customer behavior data, and a Hybrid Recommender that incorporates content based and collaborative filtering techniques

Working in a hands-on learning environment, led by our Recommendation Systems expert instructor, students will learn about and explore:

* Build industry-standard recommender systems
* Only familiarity with Python is required
* No need to wade through complicated machine learning theory to use this course.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get to grips with the different kinds of recommender systems
* Master data-wrangling techniques using the pandas library
* Building an IMDB Top 250 Clone
* Build a content-based engine to recommend movies based on movie metadata
* Employ data-mining techniques used in building recommenders
* Build industry-standard collaborative filters using powerful algorithms
* Building Hybrid Recommenders that incorporate content based and collaborative fltering

Audience & Pre-Requisites

This course is geared for attendees with Recommendation Systems with Python skills who wish to build recommendation systems is a familiarity with Python, and by the time you're finished, you will have a great grasp of how recommenders work and be in a strong position to apply the techniques that you will learn to your own problem domains.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Getting Started with Recommender Systems**

* Technical requirements
* What is a recommender system?
* Types of recommender systems

1. **Manipulating Data with the Pandas Library**

* Technical requirements
* Setting up the environment
* The Pandas library
* The Pandas DataFrame
* The Pandas Series

1. **Building an IMDB Top 250 Clone with Pandas**

* Technical requirements
* The simple recommender
* The knowledge-based recommender

1. **Building Content-Based Recommenders**

* Technical requirements
* Exporting the clean DataFrame
* Document vectors
* The cosine similarity score
* Plot description-based recommender
* Metadata-based recommender
* Suggestions for improvements

1. **Getting Started with Data Mining Techniques**

* Problem statement
* Similarity measures
* Clustering
* Dimensionality reduction
* Supervised learning
* Evaluation metrics

1. **Building Collaborative Filters**

* Technical requirements
* The framework
* User-based collaborative filtering
* Item-based collaborative filtering
* Model-based approaches

1. **Hybrid Recommenders**

* Technical requirements
* Introduction
* Case study – Building a hybrid model

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Natural Language Processing

**Course Snapshot**

* **Course:** Natural Language Processing
* **Duration:** 3 days
* **Skill-level**: Foundation-level Natural-Language-Processing skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to create machines that understand human language using the power of Python with its ecosystem of packages dedicated to NLP and AI.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Natural Language Processing** is your guide to building machines that can read and interpret human language. In it, you’ll use readily available Python packages to capture the meaning in text and react accordingly. The course expands traditional NLP approaches to include neural networks, modern deep learning algorithms, and generative techniques as you tackle real-world problems like extracting dates and names, composing text, and answering free-form questions**.**

Working in a hands-on learning environment, led by our Natural Language Processing expert instructor, students will learn about and explore:

* You’ll learn the applications to understand text and speech with extreme accuracy. The result? Chatbots that can imitate real people, meaningful resume-to-job matches, superb predictive search, and automatically generated document summaries—all at a low cost.
* New techniques, along with accessible tools like Keras and TensorFlow, make professional-quality NLP easier than ever before.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Some sentences in this course were written by NLP! Can you guess which ones?
* Working with Keras, TensorFlow, gensim, and scikit-learn
* Rule-based and data-based NLP
* Scalable pipelines

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wants to know applications to understand text and speech with extreme accuracy. The result? Chatbots that can imitate real people, meaningful resume-to-job matches, superb predictive search, and automatically generated document summaries—all at a low cost. New techniques, along with accessible tools like Keras and TensorFlow, make professional-quality NLP easier than ever before.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* This course requires a basic understanding of deep learning and intermediate Python skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Packets of thought (NLP overview)**

* Natural language vs. programming language
* The magic
* Practical applications
* Language through a computer’s “eyes”
* A brief overflight of hyperspace
* Word order and grammar
* A chatbot natural language pipeline
* Processing in depth
* Natural language IQ

1. **Build your vocabulary (word tokenization)**

* Challenges (a preview of stemming)
* Building your vocabulary with a tokenizer
* Sentiment

1. **Math with words (TF-IDF vectors)**

* Bag of words
* Vectorizing
* Zipf’s Law
* Topic modeling

1. **Finding meaning in word counts (semantic analysis)**

* From word counts to topic scores
* Latent semantic analysis
* Singular value decomposition
* Principal component analysis
* Latent Dirichlet allocation (LDiA)
* Distance and similarity
* Steering with feedback
* Topic vector power

1. **Baby steps with neural networks (perceptron’s and backpropagation)**

* Neural networks, the ingredient list

1. **Reasoning with word vectors (Word2vec)**

* Semantic queries and analogies
* Word vectors

1. **Getting words in order with convolutional neural networks (CNNs)**

* Learning meaning
* Toolkit
* Convolutional neural nets
* Narrow windows indeed

1. **Loopy (recurrent) neural networks (RNNs)**

* Remembering with recurrent networks
* Putting things together
* Let’s get to learning our past selves
* Hyperparameters
* Predicting

1. **Improving retention with long short-term memory networks**

* LSTM

1. **Sequence-to-sequence models and attention**

* Encoder-decoder architecture
* Assembling a sequence-to-sequence pipeline
* Training the sequence-to-sequence network
* Building a chatbot using sequence-to-sequence networks
* Enhancements
* In the real world

1. **Information extraction (named entity extraction and question answering)**

* Named entities and relations
* Regular patterns
* Information worth extracting
* Extracting relationships (relations)
* In the real world

1. **Getting chatty (dialog engines)**

* Pattern-matching approach
* Grounding
* Retrieval (search)
* Generative models
* Four-wheel drive
* Design process
* Trickery
* In the real world

1. **Scaling up (optimization, parallelization, and batch processing)**

* Too much of a good thing (data)
* Optimizing NLP algorithms
* Constant RAM algorithms
* Parallelizing your NLP computations
* Reducing the memory footprint during model training
* Gaining model insights with Tensor Board

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Relevant Search

Course Snapshot

* **Course:** Relevant Search
* **Duration:** 3 days
* **Skill-level**: Foundation-level Relevant Search skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to know how to return engaging search results to your users, helping you understand and leverage the internals of Lucene-based search engines.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Relevant Search** demystifies the subject and shows you that a search engine is a programmable relevance framework. You'll learn how to apply Elasticsearch or Solr to your business's unique ranking problems. The course demonstrates how to program relevance and how to incorporate secondary data sources, taxonomies, text analytics, and personalization. In practice, a relevance framework requires softer skills as well, such as collaborating with stakeholders to discover the right relevance requirements for your business. By the end, you?ll be able to achieve a virtuous cycle of provable, measurable relevance improvements over a search product?s lifetime.

Working in a hands-on learning environment, led by our Relevant Search expert instructor, students will learn about and explore:

* You'll learn how to apply Elasticsearch or Solr to your business's unique ranking problems.
* demonstrates how to program relevance and how to incorporate secondary data sources, taxonomies, text analytics, and personalization

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Techniques for debugging relevance
* Applying search engine features to real problems
* Using the user interface to guide searchers
* A systematic approach to relevance
* A business culture focused on improving search

Audience & Pre-Requisites

This course is designed for those who wants to know how to return engaging search results to your users, helping you understand and leverage the internals of Lucene-based search engines.

**Pre-Requisites:** Students should have familiar with

For developers trying to build smarter search with Elasticsearch or Solr.

Course Agenda / Topics

1. **The search relevance problemfree**

* Your goal: gaining the skills of a relevance engineer
* Why is search relevance so hard?
* Gaining insight from relevance research
* How do you solve relevance?
* More than technology: curation, collaboration, and feedback

1. **Search—under the hood**

* Search 101
* Search engine data structures
* Indexing content: extraction, enrichment, analysis, and indexing
* Document search and retrieval

1. **Debugging your first relevance problem**

* Applications to Solr and Elasticsearch: examples in Elasticsearch
* Our most prominent data set: TMDB
* Examples programmed in Python
* Your first search application
* Debugging query matching
* Debugging ranking
* Solved? Our work is never over!

1. **Taming tokensfree**

* Tokens as document features
* Controlling precision and recall
* Precision and recall—have your cake and eat it too
* Analysis strategies

1. **Basic multifield search**

* Signals and signal modeling
* TMDB—search, the final frontier!
* Signal modeling in field-centric search

1. **Term-centric search**

* What is term-centric search?
* Why do you need term-centric search?
* Performing your first term-centric searches
* Solving signal discordance in term-centric search
* Combining field-centric and term-centric strategies: having your - cake and eating it too

1. **Shaping the relevance function**

* What do we mean by score shaping?
* Boosting: shaping by promoting results
* Filtering: shaping by excluding results
* Score-shaping strategies for satisfying business needs

1. **Providing relevance feedback**

* Relevance feedback at the search box
* Relevance feedback while browsing
* Relevance feedback in the search results listing

1. **Designing a relevance-focused search application**

* Yowl! The awesome new start-up!
* Gathering information and requirements
* Designing the search application
* Deploying, monitoring, and improving
* Knowing when good is good enough

1. **The relevance-centered enterprise**

* Feedback: the bedrock of the relevance-centered enterprise
* Why user-focused culture before data-driven culture?
* Flying relevance-blind
* Relevance feedback awakenings: domain experts and expert users
* Relevance feedback maturing: content curation
* Relevance streamlined: engineer/curator pairing
* Relevance accelerated: test-driven relevance
* Beyond test-driven relevance: learning to rank

1. **Semantic and personalized search**

* Personalizing search based on user profiles
* Personalizing search based on user behavior
* information back to the search index
* Basic methods for building concept search
* Building concept search using machine learning
* The personalized search—concept search connection
* Recommendation as a generalization of search
* Best wishes on your search relevance journey

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Think Like a Data Scientist

Course Snapshot

* **Course:** Think Like a Data Scientist
* **Duration:** 3 days
* **Skill-level**: Foundation-level Data Science skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants step-by-step approach to data science, combining analytic, programming, and business perspectives into easy-to-digest techniques and thought processes for solving real world data-centric problems.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Think Like a Data Scientist** teaches you a step-by-step approach to solving real-world data-centric problems. By breaking down carefully crafted examples, you'll learn to combine analytic, programming, and business perspectives into a repeatable process for extracting real knowledge from data. As you read, you'll discover (or remember) valuable statistical techniques and explore powerful data science software. More importantly, you'll put this knowledge together using a structured process for data science. When you've finished, you'll have a strong foundation for a lifetime of data science learning and practice.

Working in a hands-on learning environment, led by our Data Science expert instructor, students will learn about and explore:

* you'll learn to combine analytic, programming, and business perspectives into a repeatable process for extracting real knowledge from data.
* you'll discover (or remember) valuable statistical techniques and explore powerful data science software.
* you'll put this knowledge together using a structured process for data science.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* The data science process, step-by-step
* How to anticipate problems
* Dealing with uncertainty
* Best practices in software and scientific thinking

Audience & Pre-Requisites

This course is designed for those who wants step-by-step approach to data science, combining analytic, programming, and business perspectives into easy-to-digest techniques and thought processes for solving real world data-centric problems.

**Pre-Requisites:** Students should have familiar with

programming skills

knowledge of basic statistics

Course Agenda / Topics

1. **Philosophies of data science**

* Data science and this course
* Awareness is valuable
* Developer vs. data scientist
* Do I need to be a software developer?
* Do I need to know statistics?
* Priorities: knowledge first, technology second, opinions third
* Best practices

1. **Setting goals by asking good questions**

* Listening to the customer
* Ask good questions—of the data
* Answering the question using data
* Setting goals
* Planning: be flexible

1. **Data all around us: the virtual wilderness**

* Data as the object of study
* Where data might live, and how to interact with it
* Scouting for data
* Example: microRNA and gene expression

1. **Data wrangling: from capture to domestication**

* Case study: best all-time performances in track and field
* Getting ready to wrangle
* Techniques and tools
* Common pitfalls

1. **Data assessment: poking and** **prodding**

* Example: the Enron email data set
* Descriptive statistics
* Check assumptions about the data
* Looking for something specific
* Rough statistical analysis

1. **Developing a plan**

* What have you learned?
* Reconsidering expectations and goals
* Planning
* Communicating new goals

1. **Statistics and modeling: concepts and foundations**

* How I think about statistics
* Statistics: the field as it relates to data science
* Mathematics
* Statistical modeling and inference
* Miscellaneous statistical methods

1. **Software: statistics**

* Spreadsheets and GUI-based applications
* Programming
* Choosing statistical software tools
* Translating statistics into software

1. **Supplementary software: bigger,** **faster, more efficient**

* Databases
* High-performance computing
* Cloud services
* Big data technologies
* Anything as a service

1. **Plan execution: putting it all together**

* Tips for executing the plan
* Modifying the plan in progress
* Results: knowing when they’re good enough
* Case study: protocols for measurement of gene activity

1. **Delivering a product**

* Understanding your customer
* Delivery media
* Content
* Example: analyzing video game play

1. **After product delivery: problems and revisions**

* Problems with the product and its use
* Feedback
* Product revisions

1. **Wrapping up: putting the project away**

* Putting the project away neatly
* Learning from the project
* Looking toward the future

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Real-World Machine Learning.

**Course Snapshot**

* **Course:** Real-World Machine Learning
* **Duration:** 3 days
* **Skill-level**: Foundation-level machine learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are to work developers the art of ML project execution. Without overdosing you on academic theory and complex mathematics
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Real-World Machine Learning** will teach you the concepts and techniques you need to be a successful machine learning practitioner without overdosing you on abstract theory and complex mathematics. By working through immediately relevant examples in Python, you’ll build skills in data acquisition and modeling, classification, and regression. You’ll also explore the most important tasks like model validation, optimization, scalability, and real-time streaming. When you’re done, you’ll be ready to successfully build, deploy, and maintain your own powerful ML systems.

Working in a hands-on learning environment, led by our Real-World Machine Learning expert instructor, students will learn about and explore:

* You’ll also explore the most important tasks like model validation, optimization, scalability, and real-time streaming
* you’ll be ready to successfully build, deploy, and maintain your own powerful ML systems.
* It introduces the day-to-day practice of machine learning, preparing you to successfully build and deploy powerful ML systems.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Predicting future behavior
* Performance evaluation and optimization
* Analyzing sentiment and making recommendations

**Audience & Pre-Requisites**

This course is geared introduces the day-to-day practice of machine learning, preparing you to successfully build and deploy powerful ML systems.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* No prior machine learning experience assumed. Readers should know Python.
* Good foundational mathematics or logic skills
* No machine learning experience or advanced math skills necessary.
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **What is machine learning?** Understanding how machines learn

* Using data to make decisions
* Following the ML workflow: from data to deployment
* Boosting model performance with advanced techniques

1. **Real-world data**

* Getting started: data collection
* Preprocessing the data for modeling
* Using data visualization

1. **Modeling and prediction**

* Basic machine-learning modeling
* Classification: predicting into buckets
* Regression: predicting numerical values

1. **Model evaluation and optimization**

* Model generalization: assessing predictive accuracy for new data
* Evaluation of classification models
* Evaluation of regression models
* Model optimization through parameter tuning

1. **Basic feature engineering**

* Motivation: why is feature engineering useful?
* Basic feature-engineering processes
* Feature selection

1. **Example: NYC taxi data**

* Data: NYC taxi trip and fare information
* Modeling

1. **Advanced feature engineering**

* Advanced text features
* Image features
* Time-series features

1. **Advanced NLP example: movie review sentiment**

* Exploring the data and use case
* Extracting basic NLP features and building the initial model
* Advanced algorithms and model deployment **considerations**

1. **Scaling machine-learning workflows**

* Before scaling up
* Scaling ML modeling pipelines
* Scaling predictions

1. **Example: digital display advertising**

* Display advertising
* Digital advertising data
* Feature engineering and modeling strategy
* Size and shape of the data
* Singular value decomposition
* Resource estimation and optimization
* Modeling
* K-nearest neighbors
* Random forests
* Other real-world considerations

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# TensorFlow Machine Learning Projects

**Course Snapshot**

* **Course:** TensorFlow Machine Learning Projects
* **Duration:** 4 days
* **Skill-level**: Foundation-level TensorFlow Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others wants to Implement TensorFlow's offerings such as TensorBoard, TensorFlow.js, TensorFlow Probability, and TensorFlow Lite to build smart automation projects
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

TensorFlow has transformed the way machine learning is perceived. TensorFlow Machine Learning Projects teaches you how to exploit the benefits—simplicity, efficiency, and flexibility—of using TensorFlow in various real-world projects. With the help of this course, you’ll not only learn how to build advanced projects using different datasets but also be able to tackle common challenges using a range of libraries from the TensorFlow ecosystem. To start with, you’ll get to grips with using TensorFlow for machine learning projects; you’ll explore a wide range of projects using TensorForest and TensorBoard for detecting exoplanets, TensorFlow.js for sentiment analysis, and TensorFlow Lite for digit classification. As you make your way through the course, you’ll build projects in various real-world domains, incorporating natural language processing (NLP), the Gaussian process, autoencoders, recommender systems, and Bayesian neural networks, along with trending areas such as Generative Adversarial Networks (GANs), capsule networks, and reinforcement learning. You’ll learn how to use the TensorFlow on Spark API and GPU-accelerated computing with TensorFlow to detect objects, followed by how to train and develop a recurrent neural network (RNN) model to generate book scripts. By the end of this course, you’ll have gained the required expertise to build full-fledged machine learning projects at work.

Working in a hands-on learning environment, led by our TensorFlow Machine Learning expert instructor, students will learn about and explore:

* Use machine learning and deep learning principles to build real-world projects
* Get to grips with TensorFlow's impressive range of module offerings
* Implement projects on GANs, reinforcement learning, and capsule network

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the TensorFlow ecosystem using various datasets and techniques
* Create recommendation systems for quality product recommendations
* Build projects using CNNs, NLP, and Bayesian neural networks
* Play Pac-Man using deep reinforcement learning
* Deploy scalable TensorFlow-based machine learning systems
* Generate your own book script using RNNs

**Audience & Pre-Requisites**

This course is for those wants to Implement TensorFlow's offerings such as TensorBoard, TensorFlow.js, TensorFlow Probability, and TensorFlow Lite to build smart automation projects.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Overview of TensorFlow and Machine Learning**

* Overview of TensorFlow and Machine Learning
* What is TensorFlow?
* The TensorFlow core
* Computation graph
* Machine learning, classification, and logistic regression
* Logistic regression with TensorFlow
* Logistic regression with Keras

1. **Using Machine Learning to Detect Exoplanets in Outer Space**

* Using Machine Learning to Detect Exoplanets in Outer Space
* What is a decision tree?
* Why do we need ensembles?
* Decision tree-based ensemble methods
* Decision tree-based ensembles in TensorFlow
* Detecting exoplanets in outer space
* Building a TFBT model for exoplanet detection

1. **Sentiment Analysis in Your Browser Using TensorFlow.js**

* Sentiment Analysis in Your Browser Using TensorFlow.js
* Understanding TensorFlow.js
* Understanding Adam Optimization
* Understanding categorical cross entropy loss
* Understanding word embeddings
* Building the sentiment analysis model
* Running the model on a browser using TensorFlow.js

1. **Digit Classification Using TensorFlow Lite**

* Digit Classification Using TensorFlow Lite
* What is TensorFlow Lite?
* Classification Model Evaluation Metrics
* Classifying digits using TensorFlow Lite

1. **Speech to Text and Topic Extraction Using NLP**

* Speech to Text and Topic Extraction Using NLP
* Speech-to-text frameworks and toolkits
* Google Speech Commands Dataset
* Neural network architecture
* Training the model

1. **Predicting Stock Prices using Gaussian Process Regression**

* Predicting Stock Prices using Gaussian Process Regression
* Understanding Bayes' rule
* Introducing Bayesian inference
* Introducing Gaussian processes
* Applying GPs to stock market prediction
* Creating a stock price prediction model
* Understanding the results obtained

1. **Credit Card Fraud Detection using Autoencoders**

* Credit Card Fraud Detection using Autoencoders
* Understanding auto-encoders
* Building a fraud detection model

1. **Generating Uncertainty in Traffic Signs Classifier Using Bayesian Neural Networks**

* Generating Uncertainty in Traffic Signs Classifier Using Bayesian Neural Networks
* Understanding Bayesian deep learning
* Understanding TensorFlow probability, variational inference, and Monte Carlo methods
* Building a Bayesian neural network

1. **Generating Matching Shoe Bags from Shoe Images Using DiscoGANs**

* Generating Matching Shoe Bags from Shoe Images Using DiscoGANs
* Understanding generative models
* Understanding DiscoGANs
* Building a DiscoGAN model

1. **Classifying Clothing Images using Capsule Networks**

* Classifying Clothing Images using Capsule Networks
* Understanding the importance of capsule networks
* Understanding capsules
* The dynamic routing algorithm
* CapsNet for classifying Fashion MNIST images
* Training and testing the model
* Reconstructing sample images
* Limitations of capsule networks

1. **Making Quality Product Recommendations Using TensorFlow**

* Making Quality Product Recommendations Using TensorFlow
* Recommendation systems
* Content-based filtering
* Collaborative filtering
* Hybrid systems
* Matrix factorization
* Introducing the Retailrocket dataset
* Exploring the Retailrocket dataset
* Pre-processing the data
* The matrix factorization model for Retailrocket recommendations
* The neural network model for Retailrocket recommendations

1. **Object Detection at a Large Scale with TensorFlow**

* Object Detection at a Large Scale with TensorFlow
* Introducing Apache Spark
* Understanding distributed TensorFlow
* Learning about TensorFlowOnSpark
* Object detection using TensorFlowOnSpark and Sparkdl

1. **Generating Book Scripts Using LSTMs**

* Generating Book Scripts Using LSTMs
* Understanding recurrent neural networks
* Pre-processing the data
* Defining the model
* Training the model
* Defining and training a text-generating model
* Generating book scripts

1. **Playing Pacman Using Deep Reinforcement Learning**

* Playing Pacman Using Deep Reinforcement Learning
* Reinforcement learning
* Reinforcement learning versus supervised and unsupervised learning
* Components of Reinforcement Learning
* OpenAI Gym
* Creating a Pacman game in OpenAI Gym
* DQN for deep reinforcement learning
* Applying DQN to a game

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

**Machine Learning to Detect Phishing Websites**

**Course Snapshot**

* **Course:** Use Machine Learning to Detect Phishing Websites
* **Duration:** 1 days
* **Skill-level**: Foundation-level Use Machine Learning to Detect Phishing Websites skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to develop a machine learning model for predicting whether or not an email that contains a link to a website is a phishing website or not.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this course, you will be filling in the role of a data scientist employed by an organization’s cybersecurity manager. Lately, the employees of the organization are receiving a lot of emails containing links to phishing websites. Your task will be to develop a machine learning model for predicting whether or not an email that contains a link to a website is a phishing website or not.

Phishing attacks are considered to be one of the most common types of online security threats, and are capable of breaking into an organization’s online security so as to extract confidential information like user passwords, financial information, and so on. The Internet Crime Report 2018 presents the effects of phishing websites.

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Load and understand a tabular dataset. As a data scientist, you should be comfortable working with tabular data.
* Query the dataset for deriving interesting reports.
* Clean the dataset accordingly so that it is well-suited for a machine learning model.
* Build and train machine learning models, like Logistic Regression and Neural Networks.
* Perform hyperparameter tuning techniques, like random search.
* Provide a summary of the performance of the machine learning models.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Loading and understanding the Phishing Websites dataset
* Asking the right questions of the data to understand it better and preparing reports
* Cleaning data
* Creating a Logistic Regression classifier as a baseline model
* Analyzing the results and using random searches to find the optimal hyperparameters of the baseline model
* Summarizing the results of your findings

**Audience & Pre-Requisites**

This course is designed for developers interested in data science and for beginner data scientists

**Pre-Requisites:** Students should have familiar with

T**OOLS**

* Basics of Python and its utility functions
* Basics of pandas
* Basics of NumPy
* Basics of scikit-learn

**TECHNIQUES**

Basics of data science

**Course Agenda / Topics**

**1. Loading and Understanding the Phishing Websites Dataset**

* Knowing the Dataset
* A Quick Tour of Pandas
* Submit Your Work

**2. Further Data Investigation and Preparing Investigation Reports**

* Getting Useful Information from the Dataset
* Submit Your Work

**3. Cleaning the Class Labels and Inspecting for Missing Values**

* Cleaning the Class Labels and Inspecting for Missing Values
* Submit Your Work

**4. Training a Logistic Regression Model**

* Training a Logistic Regression Model
* A Quick Primer on Logistic Regression
* A Brief Take on Scikit-Learn
* A Continuous Approach to Splitting Points: Logistic Regression
* Submit Your Work

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Monitoring Changes in Surface Water Using Satellite Image Data

Course Snapshot

* **Course:** Monitoring Changes in Surface Water Using Satellite Image Data
* **Duration:** 2 days
* **Skill-level**: Foundation-level Monitoring Changes in Surface Water Using Satellite Image Data skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to use the Google Collaboratory (“Colab”) coding environment to access free GPU computer resources and speed up your training times.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this,you’ll fill the shoes of a data scientist at UNESCO (United Nations Educational, Scientific and Cultural Organization). Your job involves assessing long-term changes to freshwater deposits, one of humanity’s most important resources. Recently, two European Space Agency satellites have given you a massive amount of new data in the form of satellite imagery. Your task is to build a deep learning algorithm that can process this data and automatically detect water pixels in the imagery of a region. To accomplish this, you will design, implement, and evaluate a convolutional neural network model for image pixel classification, or image segmentation. Your challenges will include compiling your data, training your model, evaluating its performance, and providing a summary of your findings to your superiors. Throughout, you’ll use the Google Collaboratory (“Colab”) coding environment to access free GPU computer resources and speed up your training times.

Working in a hands-on learning environment, led by our Monitoring Changes in Surface Water Using Satellite Image Data expert instructor, students will learn about and explore:

* Accessing cloud data servers to download satellite imagery
* Manually creating your own ground truth data from imagery
* Using the VGG-JSON image annotation format
* Using Graphical Processing Unit (GPU) computation on Google Colab
* Merging imagery and performing operations on raster datasets
* Using Keras and TensorFlow for deep learning
* Evaluating model performance by comparing estimated and observed results
* Data augmentation for boosting model training
* Optimizing model performance using experimentation
* Understanding model performance metrics (such as Dice and Jaccard scores)

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Getting started
* Data Acquisition and pre processing
* Enhancing and segmenting images
* Model training and evaluation
* Model optimization
* Reporting to unesco

Audience & Pre-Requisites

**Pre-Requisites:** Students should have familiar with

**TOOLS**

* Basic Jupyter
* Intermediate NumPy
* Intermediate Matplotlib
* Basic SciPy
* Basic pandas

**TECHNIQUES**

* Intermediate Python package installation using conda and pip
* Basics of neural networks or multi-layer perceptrons
* Basic concepts in using digital imagery for environmental monitoring

Course Agenda / Topics

1. Getting started
2. Data Acquisition and pre processing
3. Enhancing and segmenting images
4. Model training and evaluation
5. Model optimization
6. Reporting to unesco

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Human Pose Estimation with Deep Neural Networks

**Course Snapshot**

* **Course:** Human Pose Estimation with Deep Neural Networks
* **Duration:** 1 days
* **Skill-level**: Foundation-level Deep Neural Networks skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to build a human pose estimation algorithm based on convolutional neural networks.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this course, you will learn about the building blocks of deep neural networks and how to use them. After this, you will be able to build basic image classification, image segmentation or key point detection algorithms yourself. You will also learn how to use and integrate more complex models, such as an object detector into your course.

The building blocks of this course are also used in many other computer vision/machine applications. Object detection, for example, is also used for face recognition/detection, autonomous driving and OCR. The same algorithms used for key point detection are also used for image segmentation, facial landmark detection or action recognition. This course will give you the basic understanding of how all these algorithms work.

Working in a hands-on learning environment, led by our Deep Neural Networks expert instructor, students will learn about and explore:

* we will build a human pose estimation algorithm based on convolutional neural networks.
* First, we use an object detector to detect a person in an image, and then build and train a convolutional neural network from scratch to detect key points of the human body. We will use Google Colab to train a model using GPU/TPUs.
* At the end of the course, the student will have an interactive demo that uses a laptop’s webcam to do human pose estimation.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Getting the Data
* Introduction to Convolutional Neural Networks
* Object Detection
* Human Key point Estimation
* Model Deployment/Inference Demo Using the Webcam

**Audience & Pre-Requisites**

This course is for intermediate Python programmers who are familiar with machine learning. Knowledge of PyTorch and NumPy will be helpful.

**Pre-Requisites:** Students should have familiar with:

**TOOLS**

* Basics of PIL
* Basics of JSON
* Basics of Matplotlib
* Intermediate PyTorch
* Intermediate NumPy

**TECHNIQUES**

* Intermediate machine learning concepts such as classification and regression
* Basics of matrix and vector operations

**Course Agenda / Topics**

1. **Getting the Data**

* Getting the Data
* Submit Your Work

1. **Introduction to Convolutional Neural Networks**

* Introduction to Convolutional Neural Networks
* Convolutional Neural Networks (CNNs)
* Structuring Deep Learning Projects and Hyperparameters Tuning
* Submit Your Work

1. **Object Detection**

* Object Detection
* Object Detection with R-CNN, SSD, and YOLO
* Submit Your Work

1. **Human Keypoint Estimation**

* Human Keypoint Estimation—Model Training
* Human Keypoint Estimation—Transfer Learning
* ResNet
* Transfer Learning
* Submit Your Work

1. **Model Deployment / Inference Demo using the Webcam**
2. Model Deployment/Inference Demo Using the Webcam
3. Submit Your Work

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Discovering Disease Outbreaks from News Headlines

**Course Snapshot**

* **Course:** Discovering Disease Outbreaks from News Headlines
* **Duration:** 1 days
* **Skill-level**: Foundation-level Discovering Disease Outbreaks from News Headlines skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants take on the role of a data scientist at the World Health Organization (WHO).
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this course, you’ll take on the role of a data scientist at the World Health Organization (WHO). The WHO is responsible for responding to international epidemics, a critical component of which involves monitoring global news headlines for signs of disease outbreaks. However, this daily deluge of news data is too huge to manually analyze. Your challenge is to pull geographic information from headlines, and determine where in the world outbreaks are occurring. Problems you will have to solve include extracting information from text using regular expressions, using the Basemap Matplotlib extension to visualize map locations for patterns indicating an epidemic, and reporting your findings to your superiors so resources can be dispatched.

Working in a hands-on learning environment, Discovering Disease Outbreaks expert instructor, students will learn about and explore:

* you’ll develop techniques for text extraction, data manipulation, clustering, interpreting algorithm outputs,
* learn to produce an actionable report. All these skills are easily transferable to a variety of data science roles in business and other organizations.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Extracting city and country name data from text using regular expressions
* Manipulating data and matching location names to geographic coordinates
* Clustering geographic coordinates with KMeans and/or DBSCAN
* Visualizing clusters on a geographic map
* Analyzing algorithm output and tuning model settings to improve results
* Sorting between clusters based on size and within clusters based on distance
* Interpreting algorithm results in the problem domain
* Summarizing findings of a data science project effectively

**Audience & Pre-Requisites**

The course is for intermediate Python programmers who know the basics of data science. To begin this course, you will need to be familiar with

**Pre-Requisites:** Students should have familiar with:

**TOOLS**

* Basics of pandas
* Basics of scikit-learn
* Basics of text extraction
* Basics of K-means and DBSCAN clustering
* Basics of Jupyter Notebook

**Course Agenda / Topics**

1. **Extracting City and Country Information from News Headlines**
2. **Finding Geographic Locations of Headlines**
3. **Clustering Headlines Based on Location**
4. **Identifying Disease Outbreaks**
5. **Presenting the Disease Outbreak Data**
6. **Bonus Follow-up ProjectThe Coronavirus Pandemic**

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Decoding Data Science Job Postings to Improve Your Resume

**Course Snapshot**

* **Course:** Decoding Data Science Job Postings to Improve Your Resume
* **Duration:** 1 days
* **Skill-level**: Foundation-level Decoding Data Science for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to aim to take online job postings and understand the data science job market by looking at the major themes (groups) of job postings using their skill requirements
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this course, you’re a budding data scientist who has created a draft of your resume. You want to apply for data science jobs, but would like to find the jobs you have the best shot at so would like to optimize your resume for a better chance at getting one of these jobs. We will be using NLP and text analytics to search for the most relevant data science jobs from online job postings and optimize our resume for the job postings. The job post HTML pages have already been web-scraped, and we will be loading them into Python and processing the text data from there. The number of job postings that were collected is large (over one thousand), so we will need to process them with data science methods using Python. We will use text similarity methods to find the most similar job postings, and also to find key skills we’re missing from our resume. We’ll summarize our findings by printing out highlights of the text results, as well as displaying plots and word clouds of the data

Working in a hands-on learning environment, led by our Data Science expert instructor, students will learn about and explore:

* Parsing webpages with the BeautifulSoup library.
* Storing and processing data with pandas DataFrames.
* Converting raw text to numeric features (TF-IDF vectors) with the sklearn (scikit-learn) library.
* Measuring text similarity with a cosine distance function from the sklearn library.
* Dimensionality reduction with singular value decomposition (SVD) using sklearn.
* K-means clustering using sklearn.
* Creating word clouds with the WordCloud library for text cluster visualization

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Collect HTML job postings and extract relevant sections for the next steps (primarily the HTML body and skill requirements – you should find where the skill requirements are from looking at some HTML pages).
* Find which jobs are most similar to our resume based on the full text of our resume and full texts of the job posts. Future steps in the project will use this group of most similar job postings.
* Understand the major themes of job postings via the skill requirements of groups of jobs.
* Find which skills our resume is missing compared with what organizations are looking for (e.g., the skill requirements from job postings).

**Audience & Pre-Requisites**

This course is for intermediate Python programmers who know basic data science techniques. To begin this liveProject, you should be familiar with the following topics:

* **TOOLS**
  + Basics of Jupyter Notebooks
  + Basics of pandas
  + Basics of scikit-learn
* **TECHNIQUES**
  + Basics of k-means clustering
  + Basics of TF-IDF

**Course Agenda / Topics**

1. **Extracting Text from Online Job Postings**

* Our first step is to take the raw HTML job postings and extract relevant information from them, such as the skill requirements for each job.

1. **Ranking Job Postings by Similarity**

* Next, we will find the jobs that are most similar to our resume using cosine similarity**.**

1. **Clustering Job Posting Skill Requirements**

* After that, we’ll use the most similar job postings to analyze what type of skills are typically asked for by clustering the skill requirements from the job postings.

1. **Finding Missing Skills from Our Resume**

* Finally, we’ll use our most similar job postings to find which skills are missing from our resume, so we can work on those skills and add them to our resume. This should give us a better shot at getting our dream data science job.

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Practical Data Science with R

**Course Snapshot**

* **Course:** Practical Data Science with R
* **Duration:** 3 days
* **Skill-level**: Foundation-level Practical Data Science skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants learn R language and its associated tools provide a straightforward way to tackle day-to-day data science tasks without a lot of academic theory or advanced mathematics..
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Practical Data Science with R shows you how to apply the R programming language and useful statistical techniques to everyday business situations. Using examples from marketing, business intelligence, and decision support, it shows you how to design experiments (such as A/B tests), build predictive models, and present results to audiences of all levels.

Working in a hands-on learning environment, led by our Data Science expert instructor, students will learn about and explore:

It explains basic principles without the theoretical mumbo-jumbo and jumps right to the real use cases

* you'll face as you collect, curate, and analyze the data crucial to the success of your business.
* You'll apply the R programming language and statistical analysis techniques to carefully explained examples based in marketing, business intelligence, and decision support.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Data science for the business professional
* Statistical analysis using the R language
* Project lifecycle, from planning to delivery
* Numerous instantly familiar use cases
* Keys to effective data presentations

**Audience & Pre-Requisites**

This course is for intermediate Business analysts and developers are increasingly collecting, curating, analyzing, and reporting on crucial business datal.

**Pre-Requisites:** Students should have familiar with:

* Readers without a background in data science.
* Some familiarity with basic statistics, R, or another scripting language is assumed.

**Course Agenda / Topics**

1. **THE DATA SCIENCE PROCESS**

* The roles in a data science project
* Stages of a data science project
* Setting expectations

1. **LOADING DATA INTO R**

* Working with data from files
* Working with relational databases

1. **EXPLORING DATA**

* Using summary statistics to spot problems
* Spotting problems using graphics and visualization

1. **MANAGING DATA**

* Cleaning data
* Sampling for modeling and validation

1. **CHOOSING AND EVALUATING MODELS**

* Mapping problems to machine learning tasks
* Evaluating models
* Validating models

1. **MEMORIZATION METHODS**

* KDD and KDD Cup 2009
* Building single-variable models
* Building models using many variables

1. **LINEAR AND LOGISTIC REGRESSION**

* Using linear regression
* Using logistic regression

1. **UNSUPERVISED METHODS**

* Cluster analysis
* Association rules

1. **Exploring advanced methods**

* Using bagging and random forests to reduce training variance
* Using generalized additive models (GAMs) to learn non-monotone relationships
* Using kernel methods to increase data separation
* Using SVMs to model complicated decision boundaries

1. **DOCUMENTATION AND DEPLOYMENT**

* The buzz dataset
* Using knitr to produce milestone documentation
* Using comments and version control for running documentation
* Deploying models

1. **PRODUCING EFFECTIVE PRESENTATIONS**

* Presenting your results to the project sponsor
* Presenting your model to end users
* Presenting your work to other data scientists

# Building Domain Specific Language Models

**Course Snapshot**

* **Course:** Building Domain Specific Language Models
* **Duration:** 1 days
* **Skill-level**: Foundation-level Building Domain Specific Language Models skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to build the foundations of any domain-specific NLP system by creating the most a robust and efficient language model.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this course, you will be taking on the role of an NLP data scientist at **Stack Exchange**, a network of question-and-answer (Q&A) websites on topics in diverse fields. Stack Exchange has over 10M registered users and is best known for its flagship websites Stack Overflow or Ask Ubuntu. You will build statistics-focused language models using gradually more complex methods. You will evaluate and apply these models to the tasks of:

* Query completion
* Larger text generation
* Sentence selection

At the end of this course, you will be able to build the foundations of any domain-specific NLP system by creating the most a robust and efficient language model.

Working in a hands-on learning environment, led by our Building Domain Specific Language Models expert instructor, students will learn about and explore:

* Starting with building n-gram language models, which will serve as a baseline for performance evaluations,
* moving on to a more complex modeling technique based on RNNs,
* finally, using state-of-the-art language model building with the AllenNLP framework. The AllenNLP framework helps you design and evaluate deep-learning models for nearly any NLP problem.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Loading and preparing the dataset
* Building and evaluating n-gram word-based language models
* Building a word-based language model using recurrent neural networks (RNNs) and word embeddings
* Building a character-based language model with AllenNLP

**Audience & Pre-Requisites**

This course is for proficient Python programmers who have experience with text-based machine learning. This course uses Python 3.7. It is recommended that you use the Anaconda distribution of Python and conda for managing the libraries.

**Pre-Requisites:** Students should have familiar with:

**TOOLS**

* Basics of NumPy
* Basics of panda’s course
* Intermediate NLTK
* Basics of creating neural networks with PyTorch, TensorFlow, or Keras

**TECHNIQUES**

* Basics of NumPy
* Basics of pandas
* Intermediate NLTK
* Basics of creating neural networks with PyTorch, TensorFlow, or Keras

**Course Agenda / Topics**

1. **Loading and Preparing the Dataset**

* Loading and Preparing the Dataset
* Regular Expressions
* Tokenization
* Submit Your Work

1. **N-gram Language Model**

* N-gram Language Model
* Building Your Vocabularywith a Tokenizer
* Submit Your Work

1. **Deep Learning Language Model**

* Deep Learning Language Model
* Deep Learning for Text and Sequences
* Sequential NLP and Memory
* Submit Your Work

1. **Character-based Language Model with AllenNLP**

* Character-based Language Model with AllenNLP
* Sequential Labeling and Language Modeling
* Submit Your Work

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# 3D Medical Image Analysis with PyTorch

Course Snapshot

* **Course:** 3D Medical Image Analysis with PyTorch
* **Duration:** 1 days
* **Skill-level**: Foundation-level 3D Medical Image Analysis with PyTorch skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to use the deep learning framework PyTorch to implement a convolutional neural network for this task, and you will train it on the given paired data.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this course, you will be filling the role of a machine learning engineer/researcher at a healthcare technology company specializing in medical imaging applications. Your team wants to process and analyze magnetic resonance (MR) images of the brain. An MR imaging system is a flexible device that can create multiple types of images based on what a physician wants to see, but not all types of images are acquired in every scan due to time constraints. Your current processing and analysis algorithms require two types of MR images, but a new set of customer data only has one of those types. However, you have access to a fairly large, preprocessed dataset of paired examples of the two types of MR images, and you decide that deep learning would best perform this type of image transformation task.

Working in a hands-on learning environment, led by our 3D Medical Image Analysis with PyTorch expert instructor, students will learn about and explore:

* How to load and process imaging data for deep learning applications
* How to build a convolutional neural network
* How to train a neural network for a regression task
* How to evaluate the predictions of your neural network
* How to handle and visualize medical imaging data

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Training and Validation Data Setup
* Datasets and Transforms
* Create Your Neural Network
* Train the Network
* Evaluate the Results

Audience & Pre-Requisites

This course is for experienced Python programmers, familiar with object-oriented programming techniques and Python scientific computing packages. You will need to know the basics of machine learning and statistics, but this course will teach you the advanced techniques. Throughout, you’ll use the Google Collaboratory (Colab) coding environment to access free GPU computer resources and speed up your training times.

**Pre-Requisites:** Students should have familiar with:

**TOOLS**

* Basics of Matplotlib
* Basics of Jupyter Notebook
* Basics of Git
* Intermediate PyTorch

**TECHNIQUES**

* Basics of gradient descent and SGD
* Basics of Loss functions
* Basics of Back-propagation
* Basics of neural networks
* Basics of advanced functions for ANNs such as softmax, sigmoid, ReLu

Course Agenda / Topics

1. Training and Validation Data Setup

* Training and Validation Data Setup
* Volumetric Data
* Submit Your Work

1. Datasets and Transforms

* Datasets and Transforms
* Submit Your Work

1. Create Your Neural Network

* Create Your Neural Network
* Using Convolutions to Generalize
* Submit Your Work

1. Train the Network

* Train the Network
* The Mechanics of Learning
* Submit Your Work

1. Evaluate the Results

* Evaluate the Results
* Structuring Deep Learning Projects and Hyperparameters tuning
* Submit Your Work

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Growth Hacking with NLP and Sentiment Analysis

Course Snapshot

* **Course:** Growth Hacking with NLP and Sentiment Analysis
* **Duration:** 2 days
* **Skill-level**: Foundation-level 3D Growth Hacking with NLP and Sentiment Analysis skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to step into the role of a Natural Language Processing Specialist working in the Growth Hacking Team of a new video game startup.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this course, you’ll step into the role of a Natural Language Processing Specialist working in the Growth Hacking Team of a new video game startup. Your team wants to massively accelerate your company’s early growth by acquiring huge numbers of customers at the lowest possible cost. To help tailor marketing messages, your boss has asked you to map the market and find out how customers evaluate your competitors’ products. Your challenge is to create a sentiment analyzer that will give a deeper understanding of customer feedback and opinions. To do this, you’ll need to download and create a dataset from Amazon reviews, build an algorithm that will determine whether a review is positive or negative, evaluate your analyzer's performance against star ratings, and lay out your findings in a report for your manager.

Working in a hands-on learning environment, led by our Growth Hacking with NLP and Sentiment Analysis expert instructor, students will learn about and explore:

* Creating a data corpus from text reviews
* Sampling from imbalanced data
* Finding sentiment value using NLTK and dictionary-based sentiment analysis tools
* Data evaluation with scikit-learn
* Analyzing reviews using PyTorch and deep learning
* Comparing classifier performance
* Transformers-based language models
* Visualizing findings and presenting a formal report

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* CREATING YOUR DATASET
* CREATING A DICTIONARY-BASED SENTIMENT ANALYZER
* EVALUATING YOUR DICTIONARY-BASED SENTIMENT ANALYZER
* CREATING NEURAL NETWORK BASED SENTIMENT ANALYZERS
* FINDING KEY PHRASES AND WRITING A REPORT

Audience & Pre-Requisites

This course is for intermediate Python programmers who are familiar with data science. You will need to know the basics of statistics and machine learning. Previous encounters with NLP, neural networks, and PyTorch will be useful, but not essential. You’ll use the Google Collaboratory (Colab) environment for this project to access a free cloud-based GPU. To get the most out of the project, you should be familiar with.

**Pre-Requisites:** Students should have familiar with:

**TOOLS**

* Python standard library
* Basics of pandas
* Basics of Jupyter Notebook
* Basics of Colab
* Basics of scikit-learn

**TECHNIQUES**

* Basics of machine learning
* Basics of neural networks

Course Agenda / Topics

1. **CREATING YOUR DATASET**

* Creating your dataset
* Analyzing Tables Using Pandas
* Running Random Simulations in NumPy
* Introducing Annotation

1. **CREATING A DICTIONARY-BASED SENTIMENT ANALYZER**

* Creating a dictionary-based sentiment analyzer
* Build Your Vocabulary (Word Tokenization)

1. **EVALUATING YOUR DICTIONARY-BASED SENTIMENT ANALYZER**

* Evaluating your dictionary-based sentiment analyzer
* Model evaluation and optimization

1. **CREATING NEURAL NETWORK BASED SENTIMENT ANALYZERS**

* Creating neural network-based sentiment analyzers
* Introducing Deep Learning and the PyTorch Library
* Model optimization through parameter tuning
* Introducing NLP in practice: spam filtering
* What is transfer learning?

1. **FINDING KEY PHRASES AND** **WRITING A REPORT**

* Finding key phrases and writing a report

1. **SUMMARY**

* Project Conclusions

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Collective Intelligence

Course Snapshot

* **Course:** Collective Intelligence
* **Duration:** 3 days
* **Skill-level**: Foundation-level Collective Intelligence skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to develop software that you can embed in your own applications.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In the Web 2.0 era, leveraging the collective power of user contributions, interactions, and feedback is the key to market dominance. A new category of powerful programming techniques lets you discover the patterns, inter-relationships, and individual profiles—the collective intelligence—locked in the data people leave behind as they surf websites, post blogs, and interact with other users.

Collective Intelligence is a course for implementing collective-intelligence concepts using Java. It is the first Java-based course to emphasize the underlying algorithms and technical implementation of vital data gathering and mining techniques like analyzing trends, discovering relationships, and making predictions. It provides a pragmatic approach to personalization by combining content-based analysis with collaborative approaches.

Working in a hands-on learning environment, led by our Collective Intelligence expert instructor, students will learn about and explore:

* Architecture for embedding intelligence in your application
* Developing metadata about the user and content
* Gather intelligence from tagging and build tag clouds
* Introduction to intelligent web crawling and Nutch
* Harvesting information from the blogosphere
* Build a text analysis toolkit leveraging Lucene
* Business intelligence and data mining for recommendations and promotions
* Leveraging open-source data mining toolkit WEKA and the Java Data Mining (JDM) standard
* Incorporating intelligent search in your application
* Building a recommendation engine—finding related users and content
* Real-world case studies of Amazon, Google News, and Netflix personalization.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understanding collective intelligence free
* Learning from user interactions free
* Extracting intelligence from tags
* Extracting intelligence from content
* Searching the blogosphere
* Intelligent web crawling
* Data mining: process, toolkits, and standards
* Building a text analysis toolkit
* Discovering patterns with clustering
* Making predictions
* Applying intelligence in your application
* Intelligent search
* Building a recommendation engine

Audience & Pre-Requisites

**Pre-Requisites:** Students should have familiar with

* you have a basic level of Java coding skills

Course Agenda / Topics

1. Understanding collective intelligencefree

* What is collective intelligence?
* CI in web applications
* Classifying intelligence

1. Learning from user interactionsfree

* Architecture for applying intelligence
* Basics of algorithms for applying CI
* Forms of user interaction
* Converting user interaction into collective intelligence

1. Extracting intelligence from tags

* Introduction to tagging
* How to leverage tags
* Extracting intelligence from user tagging: an example
* Scalable persistence architecture for tagging
* Building tag clouds
* Finding similar tags

1. Extracting intelligence from content

* Content types and integration
* The main CI-related content types
* Extracting intelligence step by step
* Simple and composite content types

1. Searching the blogosphere

* Introducing the blogosphere
* Building a framework to search the blogosphere
* Implementing the base classes
* Integrating Technorati
* Integrating Bloglines
* Integrating providers using RSS

1. Intelligent web crawling

* Introducing web crawling
* Building an intelligent crawler step by step
* Scalable crawling with Nutch

1. Data mining: process, toolkits, and standards

* Core concepts of data mining
* Using an open source data mining framework: WEKA
* Standard data mining API: Java Data Mining (JDM)

1. Building a text analysis toolkit

* Building the text analyzers
* Building the text analysis infrastructure
* Use cases for applying the framework

1. Discovering patterns with clustering

* Clustering blog entries
* Leveraging WEKA for clustering
* Clustering using the JDM APIs

1. Making predictions

* Classification fundamentals
* Classifying blog entries using WEKA APIs
* Regression fundamentals
* Regression using WEKA
* Classification and regression using JDM

1. Intelligent search

* Search fundamentals
* Indexing with Lucene
* Searching with Lucene
* Useful tools and frameworks
* Approaches to intelligent search

1. Building a recommendation engine

* Recommendation engine fundamentals
* Content-based analysis
* Collaborative filtering
* Real-world solutions

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Scikit-learn Cookbook

Course Snapshot

* **Course:** Scikit-learn Cookbook
* **Duration:** 3 days
* **Skill-level**: Foundation-level Scikit-learn Cookbook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Learn to use scikit-learn operations and functions for Machine Learning and deep learning applications.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Python is quickly becoming the go-to language for analysts and data scientists due to its simplicity and flexibility, and within the Python data space, scikit-learn is the unequivocal choice for machine learning. This course includes walk throughs and solutions to the common as well as the not-so-common problems in machine learning, and how scikit-learn can be leveraged to perform various machine learning tasks effectively. The edition begins with taking you through recipes on evaluating the statistical properties of data and generates synthetic data for machine learning modelling. As you progress through the lessons, you will comes across recipes that will teach you to implement techniques like data pre-processing, linear regression, logistic regression, K-NN, Naïve Bayes, classification, decision trees, Ensembles and much more. Furthermore, you’ll learn to optimize your models with multi-class classification, cross validation, model evaluation and dive deeper in to implementing deep learning with scikit-learn. Along with covering the enhanced features on model section, API and new features like classifiers, regressors and estimators the lesson also contains recipes on evaluating and fine-tuning the performance of your model. By the end of this course, you will have explored plethora of features offered by scikit-learn for Python to solve any machine learning problem you come across.

Working in a hands-on learning environment, led by our Scikit-learn Cookbook expert instructor, students will learn about and explore:

* Handle a variety of machine learning tasks effortlessly by leveraging the power of scikit-learn
* Perform supervised and unsupervised learning with ease, and evaluate the performance of your model
* Practical, easy to understand recipes aimed at helping you choose the right machine learning algorithm

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Build predictive models in minutes by using scikit-learn
* Understand the differences and relationships between Classification and Regression, two types of Supervised Learning.
* Use distance metrics to predict in Clustering, a type of Unsupervised Learning
* Find points with similar characteristics with Nearest Neighbors.
* Use automation and cross-validation to find a best model and focus on it for a data product
* Choose among the best algorithm of many or use them together in an ensemble.
* Create your own estimator with the simple syntax of sklearn
* Explore the feed-forward neural networks available in scikit-learn

Audience & Pre-Requisites

This course is designed for developers wants to use scikit-learn operations and functions for Machine Learning and deep learning applications.

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **High-Performance Machine Learning – NumPy**

* High-Performance Machine Learning – NumPy
* Introduction
* NumPy basics
* Loading the iris dataset
* Viewing the iris dataset
* Viewing the iris dataset with Pandas
* Plotting with NumPy and matplotlib
* A minimal machine learning recipe – SVM classification
* Introducing cross-validation
* Putting it all together
* Machine learning overview – classification versus regression

1. **Pre-Model Workflow and Pre-Processing**

* Pre-Model Workflow and Pre-Processing
* Introduction
* Creating sample data for toy analysis
* Scaling data to the standard normal distribution
* Creating binary features through thresholding
* Working with categorical variables
* Imputing missing values through various strategies
* A linear model in the presence of outliers
* Putting it all together with pipelines
* Using Gaussian processes for regression
* Using SGD for regression

1. **Dimensionality Reduction**

* Dimensionality Reduction
* Introduction
* Reducing dimensionality with PCA
* Using factor analysis for decomposition
* Using kernel PCA for nonlinear dimensionality reduction
* Using truncated SVD to reduce dimensionality
* Using decomposition to classify with DictionaryLearning
* Doing dimensionality reduction with manifolds – t-SNE
* Testing methods to reduce dimensionality with pipelines

1. **Linear Models with scikit-learn**

* Linear Models with scikit-learn
* Introduction
* Fitting a line through data
* Fitting a line through data with machine learning
* Evaluating the linear regression model
* Using ridge regression to overcome linear regression's shortfalls
* Optimizing the ridge regression parameter
* Using sparsity to regularize models
* Taking a more fundamental approach to regularization with LARS
* References

1. **Linear Models – Logistic Regression**

* Linear Models – Logistic Regression
* Introduction
* Loading data from the UCI repository
* Viewing the Pima Indians diabetes dataset with pandas
* Looking at the UCI Pima Indians dataset web page
* Machine learning with logistic regression
* Examining logistic regression errors with a confusion matrix
* Varying the classification threshold in logistic regression
* Receiver operating characteristic – ROC analysis
* Plotting an ROC curve without context
* Putting it all together – UCI breast cancer dataset

1. **Building Models with Distance Metrics**

* Building Models with Distance Metrics
* Introduction
* Using k-means to cluster data
* Optimizing the number of centroids
* Assessing cluster correctness
* Using MiniBatch k-means to handle more data
* Quantizing an image with k-means clustering
* Finding the closest object in the feature space
* Probabilistic clustering with Gaussian mixture models
* Using k-means for outlier detection
* Using KNN for regression

1. **Cross-Validation and Post-Model Workflow**

* Cross-Validation and Post-Model Workflow
* Introduction
* Selecting a model with cross-validation
* K-fold cross validation
* Balanced cross-validation
* Cross-validation with ShuffleSplit
* Time series cross-validation
* Grid search with scikit-learn
* Randomized search with scikit-learn
* Classification metrics
* Regression metrics
* Clustering metrics
* Using dummy estimators to compare results
* Feature selection
* Feature selection on L1 norms
* Persisting models with joblib or pickle

1. **Support Vector Machines**

* Support Vector Machines
* Introduction
* Classifying data with a linear SVM
* Optimizing an SVM
* Multiclass classification with SVM
* Support vector regression

1. **Tree Algorithms and Ensembles**

* Tree Algorithms and Ensembles
* Introduction
* Doing basic classifications with decision trees
* Visualizing a decision tree with pydot
* Tuning a decision tree
* Using decision trees for regression
* Reducing overfitting with cross-validation
* Implementing random forest regression
* Bagging regression with nearest neighbors
* Tuning gradient boosting trees
* Tuning an AdaBoost regressor
* Writing a stacking aggregator with scikit-learn

1. **Text and Multiclass Classification with scikit-learn**

* Text and Multiclass Classification with scikit-learn
* Using LDA for classification
* Working with QDA – a nonlinear LDA
* Using SGD for classification
* Classifying documents with Naive Bayes
* Label propagation with semi-supervised learning

1. **Neural Networks**

* Neural Networks
* Introduction
* Perceptron classifier
* Neural network – multilayer perceptron
* Stacking with a neural network

1. **Create a Simple Estimator**

* Create a Simple Estimator
* Introduction
* Create a simple estimator

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Building Machine Learning Systems with Python

Course Snapshot

* **Course:** Building Machine Learning Systems with Python
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to get more from your data by creating practical machine learning systems with Python.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning allows systems to learn things without being explicitly programmed to do so. Python is one of the most popular languages used to develop machine learning applications, which take advantage of its extensive library support. This third edition of Building Machine Learning Systems with Python addresses recent developments in the field by covering the most-used datasets and libraries to help you build practical machine learning systems. Using machine learning to gain deeper insights from data is a key skill required by modern application developers and analysts alike. Python, being a dynamic language, allows for fast exploration and experimentation. This lesson shows you exactly how to find patterns in your raw data. You will start by brushing up on your Python machine learning knowledge and being introduced to libraries. You'll quickly get to grips with serious, real-world projects on datasets, using modeling and creating recommendation systems. With Building Machine Learning Systems with Python, you’ll gain the tools and understanding required to build your own systems, all tailored to solve real-world data analysis problems. By the end of this lesson, you will be able to build machine learning systems using techniques and methodologies such as classification, sentiment analysis, computer vision, reinforcement learning, and neural networks.

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Develop your own Python-based machine learning system
* Discover how Python offers multiple algorithms for modern machine learning systems
* Explore key Python machine learning libraries to implement in your projects

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Build a classification system that can be applied to text, images, and sound
* Employ Amazon Web Services (AWS) to run analysis on the cloud
* Solve problems related to regression using scikit-learn and TensorFlow
* Recommend products to users based on their past purchases
* Understand different ways to apply deep neural networks on structured data
* Address recent developments in the field of computer vision and reinforcement learning

Audience & Pre-Requisites

This course is designed for developers wants to get more from your data by creating practical machine learning systems with Python.

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Getting Started with Python Machine Learning**

* Getting Started with Python Machine Learning
* Machine learning and Python – a dream team

1. **Classifying with Real-World Examples**

* Classifying with Real-World Examples
* The Iris dataset
* Evaluation – holding out data and cross-validation
* How to measure and compare classifiers
* A more complex dataset and the nearest-neighbor classifier
* Which classifier to use

1. **Regression**

* Regression
* Predicting house prices with regression
* Multidimensional regression
* Cross-validation for regression
* Using Lasso or ElasticNet in scikit-learn
* Regression with TensorFlow

1. **Classification I – Detecting Poor Answers**

* Classification I – Detecting Poor Answers
* Sketching our roadmap
* Learning to classify classy answers
* Fetching the data
* Creating our first classifier
* Deciding how to improve the performance
* Using logistic regression
* Looking behind accuracy – precision and recall
* Slimming the classifier
* Ship it!
* Classification using Tensorflow

1. **Dimensionality Reduction**

* Dimensionality Reduction
* Sketching our roadmap
* Selecting features
* Feature projection
* Multidimensional scaling
* Autoencoders, or neural networks for dimensionality reduction

1. **Clustering – Finding Related Posts**

* Clustering – Finding Related Posts
* Measuring the relatedness of posts
* Preprocessing – similarity measured as a similar number of common words
* Clustering
* Solving our initial challenge
* Tweaking the parameters

1. **Recommendations**

* Recommendations
* Rating predictions and recommendations
* Splitting into training and testing
* Normalizing the training data
* A neighborhood approach to recommendations
* A regression approach to recommendations
* Combining multiple methods
* Basket analysis
* Association rule mining

1. **Artificial Neural Networks and Deep Learning**

* Artificial Neural Networks and Deep Learning
* Using TensorFlow
* Saving and restoring neural networks
* LSTM for predicting text
* LSTM for image processing

1. **Classification II – Sentiment Analysis**

* Classification II – Sentiment Analysis
* Sketching our roadmap
* Fetching the Twitter data
* Introducing the Naïve Bayes classifier
* Creating our first classifier and tuning it
* Cleaning tweets
* Taking the word types into account

1. **Topic Modeling**

* Topic Modeling
* Latent Dirichlet allocation

1. **Classification III – Music Genre Classification**

* Classification III – Music Genre Classification
* Sketching our roadmap
* Fetching the music data
* Looking at music
* Using FFT to build our first classifier
* Improving classification performance with mel frequency cepstral coefficients
* Music classification using Tensorflow

1. **Computer Vision**

* Computer Vision
* Introducing image processing
* Basic image classification
* Computing features from images
* Writing your own features
* Using features to find similar images
* Classifying a harder dataset
* Local feature representations
* Image generation with adversarial networks

1. **Reinforcement Learning**

* Reinforcement Learning
* Types of reinforcement learning
* Excelling at games

1. **Bigger Data**

* Bigger Data
* Learning about big data
* Looking under the hood
* Using jug for data analysis
* Reusing partial results
* Using Amazon Web Services
* Creating your first virtual machines
* Installing Python packages on Amazon Linux
* Running jug on our cloud machine
* Automating the generation of clusters with cfncluster

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# SAS for Finance

Course Snapshot

* **Course:** SAS for Finance
* **Duration:** 2 days
* **Skill-level**: Foundation-level SAS for Finance skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to leverage the analytical power of SAS to perform financial analysis efficiently.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

SAS is a groundbreaking tool for advanced predictive and statistical analytics used by top banks and financial corporations to establish insights from their financial data. SAS for Finance offers you the opportunity to leverage the power of SAS analytics in redefining your data. Packed with real-world examples from leading financial institutions, the author discusses statistical models using time series data to resolve business issues. This lesson shows you how to exploit the capabilities of this high-powered package to create clean, accurate financial models. You can easily assess the pros and cons of models to suit your unique business needs. By the end of this lesson, you will be able to leverage the true power of SAS to design and develop accurate analytical models to gain deeper insights into your financial data.

Working in a hands-on learning environment, led by our SAS for Finance expert instructor, students will learn about and explore:

* Leverage the power of SAS to analyze financial data with ease
* Find hidden patterns in your data, predict future trends, and optimize risk management
* Learn why leading banks and financial institutions rely on SAS for financial analysis

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand time series data and its relevance in the financial industry
* Build a time series forecasting model in SAS using advanced modeling theories
* Develop models in SAS and infer using regression and Markov chains
* Forecast inﬂation by building an econometric model in SAS for your financial planning
* Manage customer loyalty by creating a survival model in SAS using various groupings
* Understand similarity analysis and clustering in SAS using time series data

Audience & Pre-Requisites

This course is designed for developers wants to leverage the analytical power of SAS to perform financial analysis efficiently.

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Time Series Modeling in the Financial Industry**

* Time Series Modeling in the Financial Industry
* Time series illustration
* The importance of time series
* Forecasting across industries
* Characteristics of time series data
* Challenges in data
* Good versus bad forecasts
* Use of time series in the financial industry

1. **Forecasting Stock Prices and Portfolio Decisions using Time Series**

* Forecasting Stock Prices and Portfolio Decisions using Time Series
* Portfolio forecasting
* A portfolio demands decisions
* Forecasting process
* Visualization of time series data
* Dealing with multicollinearity
* Role of autocorrelation
* Scoring based on PROC REG
* Recap of key terms

1. **Credit Risk Management**

* Credit Risk Management
* Risk types
* Basel norms
* Credit risk key metrics
* Aspects of credit risk management
* PD model build

1. **Budget and Demand Forecasting**

* Budget and Demand Forecasting
* The need for the Markov model
* Business problem
* Markovian model approach
* ARIMA model approach
* Markov method for imputation

1. **Inflation Forecasting for Financial Planning**

* Inflation Forecasting for Financial Planning
* What is inflation?
* Business case for forecasting inflation
* Modeling methodology

1. **Managing Customer Loyalty Using Time Series Data**

* Managing Customer Loyalty Using Time Series Data
* Advantages of survival modeling
* Key aspects of survival analysis
* Business problem

1. **Transforming Time Series – Market Basket and Clustering**

* Transforming Time Series – Market Basket and Clustering
* Market basket analysis
* Segmentation and clustering
* MBA business problem
* Data preparation for MBA
* Assumptions for MBA
* Analysis of a set size of two
* A segmentation business problem
* Segmentation overview
* Clustering methodologies
* Segmentation suitability in the current scenario
* Segmentation modeling

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.



CATEGORY Programming

# R

**Course Snapshot**

* **Course:** R
* **Duration:** 6 days
* **Skill-level**: Foundation-level programming skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants a crash course in statistics and covers elegant methods for dealing with messy and incomplete data that are difficult to analyze using traditional methods
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

R teaches you how to use the R language by presenting examples relevant to scientific, technical, and business developers. Focusing on practical solutions, the course offers a crash course in statistics, including elegant methods for dealing with messy and incomplete data. You'll also master R's extensive graphical capabilities for exploring and presenting data visually. And this expanded includes new lessons on forecasting, data mining, and dynamic report writing.

Working in a hands-on learning environment, led by our R expert instructor, students will learn about and explore:

Focusing on practical solutions, offers a crash course in statistics and covers elegant methods for dealing with messy and incomplete data that are difficult to analyze using traditional methods.

You'll also master R's extensive graphical capabilities for exploring and presenting data visually.

this expanded edition includes new lessons on time series analysis, cluster analysis, and classification methodologies, including decision trees, random forests, and support vector machines.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Complete R language tutorial
* Using R to manage, analyze, and visualize data
* Techniques for debugging programs and creating packages
* OOP in R
* Over 160 graphs

**Audience & Pre-Requisites**

This course is designed for readers who need to solve practical data analysis problems using the R language and tools.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Some background in mathematics and statistics is helpful
* no prior experience with R or computer programming is required.

**Course Agenda / Topics**

1. Introduction to Rfree audio
2. Creating a dataset
3. Getting started with graphs
4. Basic data management
5. Advanced data management
6. Basic graphs
7. Basic statistics
8. Regression
9. Analysis of variance
10. Power analysis
11. Intermediate graphsfree audio
12. Resampling statistics and bootstrapping
13. Generalized linear models
14. Principal components and factor analysis
15. Time series
16. Cluster analysis
17. Classification
18. Advanced methods for missing data
19. Advanced graphics with ggplot2
20. Advanced programming
21. Creating a package
22. Creating dynamic reports

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.



CATEGORY Artificial Intelligence

# Mahout

Course Snapshot

* **Course:** Mahout
* **Duration:** 4 days
* **Skill-level**: Foundation-Mahout skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants practical use cases and then illustrates how Mahout can be applied to solve them.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

This course covers machine learning using Apache Mahout. Based on experience with real-world applications, it introduces practical use cases and illustrates how Mahout can be applied to solve them. It places particular focus on issues of scalability and how to apply these techniques against large data sets using the Apache Hadoop framework.

Working in a hands-on learning environment, led by our Mahout expert instructor, students will learn about and explore:

* introduction to machine learning with Apache Mahout.
* Real-world examples
* the course presents practical use cases and then illustrates how Mahout can be applied to solve them

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use group data to make individual recommendations
* Find logical clusters within your data
* Filter and refine with on-the-fly classification

Audience & Pre-Requisites

This course is written for developers familiar with Java.

**Pre-Requisites:** Students should have

* Basic Programming knowledge is needed.
* Good foundational mathematics or logic skills
* No prior experience with Mahout is assumed

Course Agenda / Topics

1. **Meet Apache Mahoutfree**
2. **Introducing recommenders**

* Defining recommendation
* Running a first recommender engine
* Evaluating a recommender
* Evaluating precision and recall
* Evaluating the GroupLens data set

1. **Representing recommender data**

* Representing preference data
* In-memory DataModels
* Coping without preference values

1. **Making recommendations**

* Understanding user-based recommendation
* Exploring the user-based recommender
* Exploring similarity metrics
* Item-based recommendation
* Slope-one recommender
* New and experimental recommenders
* Comparison to other recommenders
* Comparison to model-based recommenders

1. **Taking recommenders to production**

* Analyzing example data from a dating site
* Finding an effective recommender
* Injecting domain-specific information
* Recommending to anonymous users
* Creating a web-enabled
* Updating and monitoring the recommender

1. **Distributing recommendation computations**

* Analyzing the Wikipedia data set
* Designing a distributed item-based algorithm
* Implementing a distributed algorithm with MapReduce
* Running MapReduces with Hadoop
* Pseudo-distributing a recommender
* Looking beyond first steps with recommendations

1. **Introduction to clustering**

* Clustering basics
* Measuring the similarity of items
* Hello World: running a simple clustering example
* Exploring distance measures
* Hello World again! Trying out various distance measures

1. **Representing datafree**

* Visualizing vectors
* Representing text documents as vectors
* Generating vectors from documents
* Improving quality of vectors using normalization

1. **Clustering algorithms in Mahout**

* K-means clustering
* Beyond k-means: an overview of clustering techniques
* Fuzzy k-means clustering
* Model-based clustering
* Topic modeling using latent Dirichlet allocation (LDA)

1. **Evaluating and improving clustering quality**

* Inspecting clustering output
* Analyzing clustering output
* Improving clustering quality

1. **Taking clustering to production**

* Quick-start tutorial for running clustering on Hadoop
* Tuning clustering performance
* Batch and online clustering

1. **Real-world applications of clustering**

* Finding similar users on Twitter
* Suggesting tags for artists on Last.fm
* Analyzing the Stack Overflow data set

1. **Introduction to classification**

* Why use Mahout for classification?
* The fundamentals of classification systems
* How classification works
* Work flow in a typical classification project
* Step-by-step simple classification example

1. **Training a classifier**

* Extracting features to build a Mahout classifier
* Preprocessing raw data into classifiable data
* Converting classifiable data into vectors
* Classifying the 20 newsgroups data set with SGD
* Choosing an algorithm to train the classifier
* Classifying the 20 newsgroups data with naive Bayes

1. **Evaluating and tuning a classifier**

* Classifier evaluation in Mahout
* The classifier evaluation API
* When classifiers go bad
* Tuning for better performance

1. **Deploying a classifier**

* Process for deployment in huge systems
* Determining scale and speed requirements
* Building a training pipeline for large systems
* Integrating a Mahout classifier
* Example: a Thrift-based classification server

1. **Case study: Shop It To Me**

* Why Shop It To Me chose Mahout
* General structure of the email marketing system
* Training the model
* Speeding up classification

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning on AWS

**Course Snapshot**

* **Course:** Machine Learning on AWS
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning on AWS skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Gain expertise in ML techniques with AWS to create interactive apps using SageMaker, Apache Spark, and TensorFlow.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

AWS is constantly driving new innovations that empower data scientists to explore a variety of machine learning (ML) cloud services. This course is your comprehensive reference for learning and implementing advanced ML algorithms in AWS cloud. As you go through the lessons, you’ll gain insights into how these algorithms can be trained, tuned, and deployed in AWS using Apache Spark on Elastic Map Reduce (EMR), SageMaker, and TensorFlow. While you focus on algorithms such as XGBoost, linear models, factorization machines, and deep nets, the course will also provide you with an overview of AWS as well as detailed practical applications that will help you solve real-world problems. Every practical application includes a series of companion notebooks with all the necessary code to run on AWS. In the next few lessons, you will learn to use SageMaker and EMR Notebooks to perform a range of tasks, right from smart analytics and predictive modeling through to sentiment analysis. By the end of this course, you will be equipped with the skills you need to effectively handle machine learning projects and implement and evaluate algorithms on AWS

Working in a hands-on learning environment, led by our Machine Learning on AWS instructor, students will learn about and explore:

* Build machine learning apps on Amazon Web Services (AWS) using SageMaker, Apache Spark, and TensorFlow
* Learn model optimization and understand how to scale your models using simple and secure APIs
* Develop, train, tune, and deploy neural network models to accelerate model performance in the cloud

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Manage AI workflows by using AWS cloud to deploy services that feed smart data products
* Use SageMaker services to create recommendation models
* Scale model training and deployment using Apache Spark on EMR
* Understand how to cluster big data through EMR and seamlessly integrate it with SageMaker
* Build deep learning models on AWS using TensorFlow and deploy them as services
* Enhance your apps by combining Apache Spark and Amazon SageMaker

**Audience & Pre-Requisites**

This course is geared for attendees with basic Linux and computing skills who wish to Gain expertise in ML techniques with AWS to create interactive apps using SageMaker, Apache Spark, and TensorFlow.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, Microsoft azure and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Getting Started with Machine Learning for AWS**

* Getting Started with Machine Learning for AWS
* How AWS empowers data scientists
* Identifying candidate problems that can be solved using machine learning
* Machine learning project life cycle
* Deploying models

1. **Classifying Twitter Feeds with Naive Bayes**

* Classifying Twitter Feeds with Naive Bayes
* Classification algorithms
* Naive Bayes classifier
* Classifying text with language models
* Naive Bayes – pros and cons

1. **Predicting House Value with Regression Algorithms**

* Predicting House Value with Regression Algorithms
* Predicting the price of houses
* Understanding linear regression
* Evaluating regression models
* Implementing linear regression through scikit-learn
* Implementing linear regression through Apache Spark
* Implementing linear regression through SageMaker's linear Learner
* Understanding logistic regression
* Pros and cons of linear models

1. **Predicting User Behavior with Tree-Based Methods**

* Predicting User Behavior with Tree-Based Methods
* Understanding decision trees
* Understanding random forest algorithms
* Understanding gradient boosting algorithms
* Predicting clicks on log streams

1. **Customer Segmentation Using Clustering Algorithms**

* Customer Segmentation Using Clustering Algorithms
* Understanding How Clustering Algorithms Work
* Clustering with Apache Spark on EMR

1. **Analyzing Visitor Patterns to Make Recommendations**

* Analyzing Visitor Patterns to Make Recommendations
* Making theme park attraction recommendations through Flickr data
* Collaborative filtering
* Finding recommendations through Apache Spark's ALS
* Recommending attractions through SageMaker Factorization Machines

1. **Implementing Deep Learning Algorithms**

* Implementing Deep Learning Algorithms
* Understanding deep learning
* Applications of deep learning
* Understanding deep learning algorithms
* Understanding convolutional neural networks

1. **Implementing Deep Learning with TensorFlow on AWS**

* Implementing Deep Learning with TensorFlow on AWS
* About TensorFlow
* TensorFlow as a general machine learning library
* Training and serving the TensorFlow model through SageMaker
* Creating a custom neural net with TensorFlow

1. **Image Classification and Detection with SageMaker**

* Image Classification and Detection with SageMaker
* Introducing Amazon SageMaker for image classification
* Training a deep learning model using Amazon SageMaker
* Classifying images using Amazon SageMaker

1. **Working with AWS Comprehend**

* Working with AWS Comprehend
* Introducing Amazon Comprehend
* Accessing AmazonComprehend
* Named-entity recognition using Comprehend
* Sentiment analysis using Comprehend
* Text classification using Comprehend

1. **Using AWS Rekognition**

* Using AWS Rekognition
* Introducing Amazon Rekognition
* Implementing object and scene detection
* Implementing facial analysis

1. **Building Conversational Interfaces Using AWS Lex**

* Building Conversational Interfaces Using AWS Lex
* Introducing Amazon Lex
* Building a custom chatbot using Amazon Lex

1. **Creating Clusters on AWS**

* Creating Clusters on AWS
* Choosing your instance types
* Distributed deep learning

1. **Optimizing Models in Spark and SageMaker**

* Optimizing Models in Spark and SageMaker
* The importance of model optimization
* Automatic hyperparameter tuning
* Hyperparameter tuning in Apache Spark
* Hyperparameter tuning in SageMaker

1. **Tuning Clusters for Machine Learning**

* Tuning Clusters for Machine Learning
* Introduction to the EMR architecture
* Tuning EMR for different applications
* Managing data pipelines with Glue

1. **Deploying Models Built in AWS**

* Deploying Models Built in AWS
* SageMaker model deployment
* Apache Spark model deployment

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning for Developers

**Course Snapshot**

* **Course:** Machine Learning for Developers
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants one-stop guide to becoming a Machine Learning expert.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Most of us have heard about the term Machine Learning, but surprisingly the question frequently asked by developers across the globe is, “How do I get started in Machine Learning?”. One reason could be attributed to the vastness of the subject area because people often get overwhelmed by the abstractness of ML and terms such as regression, supervised learning, probability density function, and so on. This book is a systematic guide teaching you how to implement various Machine Learning techniques and their day-to-day application and development. You will start with the very basics of data and mathematical models in easy-to-follow language that you are familiar with; you will feel at home while implementing the examples. The course will introduce you to various libraries and frameworks used in the world of Machine Learning, and then, without wasting any time, you will get to the point and implement Regression, Clustering, classification, Neural networks, and more with fun examples. As you get to grips with the techniques, you’ll learn to implement those concepts to solve real-world scenarios for ML applications such as image analysis, Natural Language processing, and anomaly detections of time series data. By the end of the course, you will have learned various ML techniques to develop more efficient and intelligent applications.

Working in a hands-on learning environment, led by our Machine Learning instructor, students will learn about and explore:

* Learn to develop efficient and intelligent applications by leveraging the power of Machine Learning
* A highly practical guide explaining the concepts of problem solving in the easiest possible manner
* Implement Machine Learning in the most practical way

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn the math and mechanics of Machine Learning via a developer-friendly approach
* Get to grips with widely used Machine Learning algorithms/techniques and how to use them to solve real problems
* Get a feel for advanced concepts, using popular programming frameworks.
* Prepare yourself and other developers for working in the new ubiquitous field of Machine Learning
* Get an overview of the most well known and powerful tools, to solve computing problems using Machine Learning.
* Get an intuitive and down-to-earth introduction to current Machine Learning areas, and apply these concepts on interesting and cutting-edge problems.

**Audience & Pre-Requisites**

This course is geared for attendees with basic Linux and computing skills who wish to Gain one-stop guide to becoming a Machine Learning expert.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, Microsoft azure and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction - Machine Learning and Statistical Science**

* Introduction - Machine Learning and Statistical Science
* Machine learning in the bigger picture
* Tools of the trade–programming language and libraries
* Basic mathematical concepts

1. **The Learning Process**

* The Learning Process
* Understanding the problem
* Dataset definition and retrieval
* Feature engineering
* Dataset preprocessing
* Model definition
* Loss function definition
* Model fitting and evaluation
* Model implementation and results interpretation

1. **Clustering**

* Clustering
* Grouping as a human activity
* Automating the clustering process
* Finding a common center - K-means
* Nearest neighbors
* K-NN sample implementation

1. **Linear and Logistic Regression**

* Linear and Logistic Regression
* Regression analysis
* Linear regression
* Data exploration and linear regression in practice
* Logistic regression

1. **Neural Networks**

* Neural Networks
* History of neural models
* Implementing a simple function with a single-layer perceptron

1. **Convolutional Neural Networks**

* Convolutional Neural Networks
* Origin of convolutional neural networks
* Deep neural networks
* Deploying a deep neural network with Keras
* Exploring a convolutional model with Quiver

1. **Recurrent Neural Networks**

* Recurrent Neural Networks
* Solving problems with order — RNNs
* LSTM
* Univariate time series prediction with energy consumption data

1. **Recent Models and Developments**

* Recent Models and Developments
* GANs
* Reinforcement learning
* Basic RL techniques: Q-learning

1. **Software Installation and Configuration**

* Software Installation and Configuration
* Linux installation
* macOS X environment installation
* Windows installation

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with IBM Watson

**Course Snapshot**

* **Course:** Machine Learning with IBM Watson
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning with IBM Watson skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Learn how to build complete machine learning systems with IBM Cloud and Watson Machine learning services.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

IBM Cloud is a collection of cloud computing services for data analytics using machine learning and artificial intelligence (AI). This course is a complete guide to help you become well versed with machine learning on the IBM Cloud using Python. Machine Learning with IBM Watson starts with supervised and unsupervised machine learning concepts, in addition to providing you with an overview of IBM Cloud and Watson Machine Learning. You'll gain insights into running various techniques, such as K-means clustering, K-nearest neighbor (KNN), and time series prediction in IBM Cloud with real-world examples. The course will then help you delve into creating a Spark pipeline in Watson Studio. You will also be guided through deep learning and neural network principles on the IBM Cloud using TensorFlow. With the help of NLP techniques, you can then brush up on building a chatbot. In later lessons, you will cover three powerful case studies, including the facial expression classification platform, the automated classification of lithofacies, and the multi-biometric identity authentication platform, helping you to become well versed with these methodologies. By the end of this course, you will be ready to build efficient machine learning solutions on the IBM Cloud and draw insights from the data at hand using real-world examples.

Working in a hands-on learning environment, led by our Machine Learning with IBM Watson instructor, students will learn about and explore:

* Implement data science and machine learning techniques to draw insights from real-world data
* Understand what IBM Cloud platform can help you to implement cognitive insights within applications
* Understand the role of data representation and feature extraction in any machine learning system

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand key characteristics of IBM machine learning services
* Run supervised and unsupervised techniques in the cloud
* Understand how to create a Spark pipeline in Watson Studio
* Implement deep learning and neural networks on the IBM Cloud with TensorFlow
* Create a complete, cloud-based facial expression classification solution
* Use biometric traits to build a cloud-based human identification system

**Audience & Pre-Requisites**

This course is geared for attendees with basic Linux and computing skills who wish to Learn how to build complete machine learning systems with IBM Cloud and Watson Machine learning services.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction to IBM Cloud**

* Introduction to IBM Cloud
* Understanding IBM Cloud
* Accessing the IBM Cloud
* Cloud resources
* The IBM Cloud and Watson Machine Learning services
* Setting up the environment
* Watson Studio Cloud
* Setting up a new project
* Data visualization tutorial

1. **Feature Extraction - A Bag of Tricks**

* Feature Extraction - A Bag of Tricks
* Preprocessing
* Dimensional reduction
* Data fusion
* A bag of tricks

1. **Supervised Machine Learning Models for Your Data**

* Supervised Machine Learning Models for Your Data
* Model selection
* Testing the model
* Classification
* Regression
* Testing the predictive capability

1. **Implementing Unsupervised Algorithms**

* Implementing Unsupervised Algorithms
* Unsupervised learning
* Semi-supervised learning
* Anomaly detection
* Online or batch learning

1. **Machine Learning Workouts on IBM Cloud**

* Machine Learning Workouts on IBM Cloud
* Watson Studio and Python
* Setting up the environment
* Data cleansing and preparation
* K-means clustering using Python
* K-nearest neighbors
* Time series prediction example

1. **Using Spark with IBM Watson Studio**

* Using Spark with IBM Watson Studio
* Introduction to Apache Spark
* Watson Studio and Spark
* Creating a Spark-enabled notebook
* Creating a Spark pipeline in Watson Studio
* Data preparation
* A data analysis and visualization example

1. **Deep Learning Using TensorFlow on the IBM Cloud**

* Deep Learning Using TensorFlow on the IBM Cloud
* Introduction to deep learning
* TensorFlow basics
* Neural networks and TensorFlow
* An example
* TensorFlow and image classifications
* Additional preparation

1. **Creating a Facial Expression Platform on IBM Cloud**

* Creating a Facial Expression Platform on IBM Cloud
* Understanding facial expression classification
* Exploring expression databases
* Preprocessing faces
* Preparing the environment
* Learning the expression classifier

1. **The Automated Classification of Lithofacies Formation Using ML**

* The Automated Classification of Lithofacies Formation Using ML
* Understanding lithofacies
* Exploring the data
* Training the classifier
* Evaluating the classifier

1. **Building a Cloud-Based Multibiometric Identity Authentication Platform**

* Building a Cloud-Based Multibiometric Identity Authentication Platform
* Understanding biometrics
* Exploring biometric data
* Feature extraction
* Multimodal fusion

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Scala for Machine Learning

**Course Snapshot**

* **Course:** Scala for Machine Learning
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants Leverage Scala and Machine Learning to study and construct systems that can learn from data.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The discovery of information through data clustering and classification is becoming a key differentiator for competitive organizations. Machine learning applications are everywhere, from self-driving cars, engineering design, logistics, manufacturing, and trading strategies, to detection of genetic anomalies. The course is your one stop guide that introduces you to the functional capabilities of the Scala programming language that are critical to the creation of machine learning algorithms such as dependency injection and implicits. You start by learning data preprocessing and filtering techniques. Following this, you'll move on to unsupervised learning techniques such as clustering and dimension reduction, followed by probabilistic graphical models such as Naïve Bayes, hidden Markov models and Monte Carlo inference. Further, it covers the discriminative algorithms such as linear, logistic regression with regularization, kernelization, support vector machines, neural networks, and deep learning. You’ll move on to evolutionary computing, multibandit algorithms, and reinforcement learning. Finally, the course includes a comprehensive overview of parallel computing in Scala and Akka followed by a description of Apache Spark and its ML library. With updated codes based on the latest version of Scala and comprehensive examples, this book will ensure that you have more than just a solid fundamental knowledge in machine learning with Scala.

Working in a hands-on learning environment, led by our Machine Learning instructor, students will learn about and explore:

* Explore a broad variety of data processing, machine learning, and genetic algorithms through diagrams, mathematical formulation, and updated source code in Scala
* Take your expertise in Scala programming to the next level by creating and customizing AI applications
* Experiment with different techniques and evaluate their benefits and limitations using real-world applications in a tutorial style

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Build dynamic workflows for scientific computing
* Leverage open source libraries to extract patterns from time series
* Write your own classification, clustering, or evolutionary algorithm
* Perform relative performance tuning and evaluation of Spark
* Master probabilistic models for sequential data
* Experiment with advanced techniques such as regularization and kernelization
* Dive into neural networks and some deep learning architecture
* Apply some basic multiarm-bandit algorithms
* Solve big data problems with Scala parallel collections, Akka actors, and Apache Spark clusters
* Apply key learning strategies to a technical analysis of financial markets

**Audience & Pre-Requisites**

This course is geared for attendees with basic Linux and computing skills who wish to Learn leverage Scala and Machine Learning to study and construct systems that can learn from data.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Getting Started**

* Getting Started
* Mathematical notations for the curious
* Why machine learning?
* Why Scala?
* Model categorization
* Taxonomy of machine learning algorithms
* Leveraging Java libraries
* Tools and frameworks
* Source code
* Let's kick the tires

1. **Data Pipelines**

* Data Pipelines
* Modeling
* Defining a methodology
* Monadic data transformation
* Workflow computational model
* Profiling data
* Assessing a model

1. **Data Preprocessing**

* Data Preprocessing
* Time series in Scala
* Moving averages
* Fourier analysis
* The discrete Kalman filter
* Alternative preprocessing techniques

1. **Unsupervised Learning**

* Unsupervised Learning
* K-mean clustering
* Expectation-Maximization (EM)

1. **Dimension Reduction**

* Dimension Reduction
* Challenging model complexity
* The divergences
* Principal components analysis (PCA)
* Nonlinear models

1. **Naïve Bayes Classifiers**

* Naïve Bayes Classifiers
* Probabilistic graphical models
* Naïve Bayes classifiers
* Multivariate Bernoulli classification
* Naïve Bayes and text mining
* Pros and cons

1. **Sequential Data Models**

* Sequential Data Models
* Markov decision processes
* The hidden Markov model (HMM)
* Conditional random fields
* Regularized CRF and text analytics
* Comparing CRF and HMM
* Performance consideration

1. **Monte Carlo Inference**

* Monte Carlo Inference
* The purpose of sampling
* Gaussian sampling
* Monte Carlo approximation
* Bootstrapping with replacement
* Markov Chain Monte Carlo (MCMC)

1. **Regression and Regularization**

* Regression and Regularization
* Linear regression
* Regularization
* Numerical optimization
* Logistic regression

1. **Multilayer Perceptron**

* Multilayer Perceptron
* Feed-forward neural networks (FFNN)
* The multilayer perceptron (MLP)
* Evaluation
* Benefits and limitations

1. **Deep Learning**

* Deep Learning
* Sparse autoencoder
* Restricted Boltzmann Machines (RBMs)
* Convolution neural network

1. **Kernel Models and SVM**

* Kernel Models and SVM
* Kernel functions
* The support vector machine (SVM)
* Performance considerations

1. **Evolutionary Computing**

* Evolutionary Computing
* Evolution
* Genetic algorithms and machine learning
* Genetic algorithm components
* Implementation
* GA for trading strategies
* Advantages and risks of genetic algorithms

1. **Multiarmed Bandits**

* Multiarmed Bandits
* K-armed bandit
* Thompson sampling
* Upper bound confidence

1. **Reinforcement Learning**

* Reinforcement Learning
* Reinforcement learning
* Learning classifier systems

1. **Parallelism in Scala and Akka**

* Parallelism in Scala and Akka
* Overview
* Scala
* Scalability with Actors
* Akka

1. **Apache Spark MLlib**

* Apache Spark MLlib
* Overview
* Apache Spark core
* MLlib library
* Reusable ML pipelines
* Extending Spark
* Streaming engine
* Performance evaluation
* Pros and cons

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Deep-Learning-for-Natural-Language-Processing

Course Snapshot

* **Course:** Deep-Learning-for-Natural-Language-Processing
* **Duration:** 2 days
* **Skill-level**: Foundation-level Deep-Learning-for-Natural-Language-Processing skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to explore the most challenging NLP issues and learn how to solve them with deep learning!
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Deep Learning for Natural Language Processing teaches you to apply state-of-the-art deep learning approaches to natural language processing tasks. You’ll learn key NLP concepts like neural word embeddings, auto-encoders, part-of-speech tagging, parsing, and semantic inference. Then you’ll dive deeper into advanced topics including deep memory-based NLP, linguistic structure, and hyperparameters for deep NLP. Along the way, you’ll pick up emerging best practices and gain hands-on experience with a myriad of examples, all written in Python and the powerful Keras library. By the time you’re done reading this invaluable course, you’ll be solving a wide variety of NLP problems with cutting-edge deep learning techniques!

Working in a hands-on learning environment, led by our Natural Language Processing expert instructor, students will learn about and explore:

* explore the most challenging NLP issues and learn how to solve them with deep learning!
* A goldmine of unstructured textual data already exists, largely untapped simply because it doesn’t follow any predefined format.
* NLP is poised to conquer that data with its impressive abilities to scan for keywords and phrases and discern sentiment and preferences.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* An overview of NLP and deep learning
* One-hot text representations
* Word embeddings
* Models for textual similarity
* Sequential NLP
* Semantic role labeling
* Deep memory-based NLP
* Linguistic structure
* Hyperparameters for deep NLP

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to learn and build NLP applications, and know exactly what to look for when approaching new challenges.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Deep learning for NLP**

* Overview of the course
* A selection of machine learning methods for NLP
* Deep Learning
* Vector representations of language
* Vector sanitization
* Wrapping up

1. **Deep learning and language: the basics**

* Basic architectures of deep learning
* Deep learning and NLP: a new paradigm
* Wrapping up

1. **Text embeddings**

* Embeddings
* From words to vectors: word2vec
* From documents to vectors: doc2vec
* Wrapping up
* External resources

1. **Textual similarity**

* The problem
* The data
* Data representation
* Models for measuring similarity
* Authorship attribution
* Authorship verification
* Wrap up

1. **Sequential NLP and memory**

* Memory and language
* Data and data processing
* Question Answering with sequential models
* Data and software resources

1. **Episodic memory for NLP**

* Memory networks for sequential NLP
* Data and data processing
* Strongly supervised memory networks: experiments and results
* Semi-supervised memory networks
* Semi-supervised memory networks: experiments and results
* Code and data

1. **Attention**

* Neural attention
* Data
* Static attention: MLP
* Temporal attention: LSTM

1. **Multitask learning**

* Introduction
* Data
* Consumer reviews: Yelp and Amazon
* Reuters topic classification
* Part-of-speech and named entity recognition data

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Ensemble Learning with R

Course Snapshot

* **Course:** Ensemble Learning with R
* **Duration:** 3 days
* **Skill-level**: Foundation-level Ensemble Learning with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Explore powerful R packages to create predictive models using ensemble methods
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Ensemble techniques are used for combining two or more similar or dissimilar machine learning algorithms to create a stronger model. Such a model delivers superior prediction power and can give your datasets a boost in accuracy. Ensemble Learning with R begins with the important statistical resampling methods. You will then walk through the central trilogy of ensemble techniques – bagging, random forest, and boosting – then you'll learn how they can be used to provide greater accuracy on large datasets using popular R packages. You will learn how to combine model predictions using different machine learning algorithms to build ensemble models. In addition to this, you will explore how to improve the performance of your ensemble models. By the end of this course, you will have learned how machine learning algorithms can be combined to reduce common problems and build simple efficient ensemble models with the help of real-world examples.

Working in a hands-on learning environment, led by our Ensemble Learning with R instructor, students will learn about and explore:

* Implement machine learning algorithms to build ensemble-efficient models
* Explore powerful R packages to create predictive models using ensemble methods
* Learn to build ensemble models on large datasets using a practical approach.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Carry out an essential review of re-sampling methods, bootstrap, and jackknife
* Explore the key ensemble methods: bagging, random forests, and boosting
* Use multiple algorithms to make strong predictive models
* Enjoy a comprehensive treatment of boosting methods
* Supplement methods with statistical tests, such as ROC
* Walk through data structures in classification, regression, survival, and time series data
* Use the supplied R code to implement ensemble methods
* Learn stacking method to combine heterogeneous machine learning models

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to Explore powerful R packages to create predictive models using ensemble methods

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Introduction to Ensemble Techniques**

* Introduction to Ensemble Techniques
* Datasets
* Statistical/machine learning models
* The right model dilemma!
* An ensemble purview
* Complementary statistical tests

1. **Bootstrapping**

* Bootstrapping
* Technical requirements
* The jackknife technique
* Bootstrap – a statistical method
* The boot package
* Bootstrap and testing hypotheses
* Bootstrapping regression models
* Bootstrapping survival models\*
* Bootstrapping time series models\*

1. **Bagging**

* Bagging
* Technical requirements
* Classification trees and pruning
* Bagging
* k-NN classifier
* k-NN bagging

1. **Random Forests**

* Random Forests
* Technical requirements
* Random Forests
* Variable importance
* Proximity plots
* Random Forest nuances
* Comparisons with bagging
* Missing data imputation
* Clustering with Random Forest

1. **The Bare Bones Boosting Algorithms**

* The Bare Bones Boosting Algorithms
* Technical requirements
* The general boosting algorithm
* Adaptive boosting
* Gradient boosting
* Using the adabag and gbm packages
* Variable importance
* Comparing bagging, random forests, and boosting

1. **Boosting Refinements**

* Boosting Refinements
* Technical requirements
* Why does boosting work?
* The gbm package
* The xgboost package
* The h2o package

1. **The General Ensemble Technique**

* The General Ensemble Technique
* Technical requirements
* Why does ensembling work?
* Ensembling by voting
* Ensembling by averaging
* Stack ensembling

1. **Ensemble Diagnostics**

* Ensemble Diagnostics
* Technical requirements
* What is ensemble diagnostics?
* Ensemble diversity
* Pairwise measure
* Interrating agreement

1. **Ensembling Regression Models**

* Ensembling Regression Models
* Technical requirements
* Pre-processing the housing data
* Visualization and variable reduction
* Regression models
* Bagging and Random Forests
* Boosting regression models
* Stacking methods for regression models

1. **Ensembling Survival Models**

* Ensembling Survival Models
* Core concepts of survival analysis
* Nonparametric inference
* Regression models – parametric and Cox proportional hazards models
* Survival tree
* Ensemble survival models

1. **Ensembling Time Series Models**

* Ensembling Time Series Models
* Technical requirements
* Time series datasets
* Time series visualization
* Core concepts and metrics
* Essential time series models
* Bagging and time series
* Ensemble time series models

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with scikit-learn

Course Snapshot

* **Course:** Machine Learning with scikit-learn
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning with scikit-learn skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Use scikit-learn to apply machine learning to real-world problems
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning is the buzzword bringing computer science and statistics together to build smart and efficient models. Using powerful algorithms and techniques offered by machine learning you can automate any analytical model. This course examines a variety of machine learning models including popular machine learning algorithms such as k-nearest neighbors, logistic regression, naive Bayes, k-means, decision trees, and artificial neural networks. It discusses data preprocessing, hyperparameter optimization, and ensemble methods. You will build systems that classify documents, recognize images, detect ads, and more. You will learn to use scikit-learn’s API to extract features from categorical variables, text and images; evaluate model performance, and develop an intuition for how to improve your model’s performance. By the end of this course, you will master all required concepts of scikit-learn to build efficient models at work to carry out advanced tasks with the practical approach.

Working in a hands-on learning environment, led by our Machine Learning with scikit-learn instructor, students will learn about and explore:

* Master popular machine learning models including k-nearest neighbors, random forests, logistic regression, k-means, naive Bayes, and artificial neural networks
* Learn how to build and evaluate performance of efficient models using scikit-learn
* Practical guide to master your basics and learn from real life applications of machine learning

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Review fundamental concepts such as bias and variance
* Extract features from categorical variables, text, and images
* Predict the values of continuous variables using linear regression and K Nearest Neighbors
* Classify documents and images using logistic regression and support vector machines
* Create ensembles of estimators using bagging and boosting techniques
* Discover hidden structures in data using K-Means clustering
* Evaluate the performance of machine learning systems in common tasks

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to Use scikit-learn to apply machine learning to real-world problems

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **The Fundamentals of Machine Learning**

* The Fundamentals of Machine Learning
* Defining machine learning
* Learning from experience
* Machine learning tasks
* Training data, testing data, and validation data
* Bias and variance
* An introduction to scikit-learn
* Installing scikit-learn
* Installing pandas, Pillow, NLTK, and matplotlib

1. **Simple Linear Regression**

* Simple Linear Regression
* Simple linear regression
* Evaluating the model

1. **Classification and Regression with k-Nearest Neighbors**

* Classification and Regression with k-Nearest Neighbors
* K-Nearest Neighbors
* Lazy learning and non-parametric models
* Classification with KNN
* Regression with KNN

1. **Feature Extraction**

* Feature Extraction
* Extracting features from categorical variables
* Standardizing features
* Extracting features from text
* Extracting features from images

1. **From Simple Linear Regression to Multiple Linear Regression**

* From Simple Linear Regression to Multiple Linear Regression
* Multiple linear regression
* Polynomial regression
* Regularization
* Applying linear regression
* Gradient descent

1. **From Linear Regression to Logistic Regression**

* From Linear Regression to Logistic Regression
* Binary classification with logistic regression
* Spam filtering
* Tuning models with grid search
* Multi-class classification
* Multi-label classification and problem transformation

1. **Naive Bayes**

* Naive Bayes
* Bayes' theorem
* Generative and discriminative models
* Naive Bayes
* Naive Bayes with scikit-learn

1. **Nonlinear Classification and Regression with Decision Trees**

* Nonlinear Classification and Regression with Decision Trees
* Decision trees
* Training decision trees
* Decision trees with scikit-learn

1. **From Decision Trees to Random Forests and Other Ensemble Methods**

* From Decision Trees to Random Forests and Other Ensemble Methods
* Bagging
* Boosting
* Stacking

1. **The Perceptron**

* The Perceptron
* The perceptron
* Limitations of the perceptron

1. **From the Perceptron to Support Vector Machines**

* From the Perceptron to Support Vector Machines
* Kernels and the kernel trick
* Maximum margin classification and support vectors
* Classifying characters in scikit-learn

1. **From the Perceptron to Artificial Neural Networks**

* From the Perceptron to Artificial Neural Networks
* Nonlinear decision boundaries
* Feed-forward and feedback ANNs
* Multi-layer perceptrons
* Training multi-layer perceptrons

1. **K-means**

* K-means
* Clustering
* K-means
* Evaluating clusters
* Image quantization
* Clustering to learn features

1. **Dimensionality Reduction with Principal Component Analysis**

* Dimensionality Reduction with Principal Component Analysis
* Principal component analysis
* Visualizing high-dimensional data with PCA
* Face recognition with PCA

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning for OpenCV

Course Snapshot

* **Course:** Machine Learning for OpenCV
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Expand OpenCV knowledge and master key concepts of machine learning using this practical, hands-on guide
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine Learning is no longer just a buzzword, it is all around us: from protecting your email, to automatically tagging friends in pictures, to predicting what movies you like. Computer vision is one of today's most exciting application fields of Machine Learning, with Deep Learning driving innovative systems such as self-driving cars and Google’s DeepMind. OpenCV lies at the intersection of these topics, providing a comprehensive open-source library for classic as well as state-of-the-art computer vision and Machine Learning algorithms. In combination with Python Anaconda, you will have access to all the open-source computing libraries you could possibly ask for. Machine Learning for OpenCV begins by introducing you to the essential concepts of statistical learning, such as classification and regression. Once all the basics are covered, you will start exploring various algorithms such as decision trees, support vector machines, and Bayesian networks, and learn how to combine them with other OpenCV functionality. As the course progresses, so will your Machine Learning skills, until you are ready to take on today's hottest topic in the field: Deep Learning. By the end of this, you will be ready to take on your own Machine Learning problems, either by building on the existing source code or developing your own algorithm from scratch!

Working in a hands-on learning environment, led by our Machine Learning instructor, students will learn about and explore:

* Load, store, edit, and visualize data using OpenCV and Python
* Grasp the fundamental concepts of classification, regression, and clustering
* Understand, perform, and experiment with Machine Learning techniques using this easy-to-follow guide
* Evaluate, compare, and choose the right algorithm for any task

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Explore and make effective use of OpenCV's Machine Learning module
* Learn deep learning for computer vision with Python
* Master linear regression and regularization techniques
* Classify objects such as flower species, handwritten digits, and pedestrians
* Explore the effective use of support vector machines, boosted decision trees, and random forests
* Get acquainted with neural networks and Deep Learning to address real-world problems
* Discover hidden structures in your data using k-means clustering
* Get to grips with data pre-processing and feature engineering

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to expand OpenCV knowledge and master key concepts of machine learning using this practical, hands-on guide

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **A Taste of Machine Learning**

* A Taste of Machine Learning
* Getting started with machine learning
* Problems that machine learning can solve
* Getting started with Python
* Getting started with OpenCV
* Installation

1. **Working with Data in OpenCV and Python**

* Working with Data in OpenCV and Python
* Understanding the machine learning workflow
* Dealing with data using OpenCV and Python

1. **First Steps in Supervised Learning**

* First Steps in Supervised Learning
* Understanding supervised learning
* Using classification models to predict class labels
* Using regression models to predict continuous outcomes
* Classifying iris species using logistic regression

1. **Representing Data and Engineering Features**

* Representing Data and Engineering Features
* Understanding feature engineering
* Preprocessing data
* Understanding dimensionality reduction
* Representing categorical variables
* Representing text features
* Representing images

1. **Using Decision Trees to Make a Medical Diagnosis**

* Using Decision Trees to Make a Medical Diagnosis
* Understanding decision trees
* Using decision trees to diagnose breast cancer
* Using decision trees for regression

1. **Detecting Pedestrians with Support Vector Machines**

* Detecting Pedestrians with Support Vector Machines
* Understanding linear support vector machines
* Dealing with nonlinear decision boundaries
* Detecting pedestrians in the wild

1. **Implementing a Spam Filter with Bayesian Learning**

* Implementing a Spam Filter with Bayesian Learning
* Understanding Bayesian inference
* Implementing your first Bayesian classifier
* Classifying emails using the naive Bayes classifier

1. **Discovering Hidden Structures with Unsupervised Learning**

* Discovering Hidden Structures with Unsupervised Learning
* Understanding unsupervised learning
* Understanding k-means clustering
* Understanding expectation-maximization
* Compressing color spaces using k-means
* Classifying handwritten digits using k-means
* Organizing clusters as a hierarchical tree

1. **Using Deep Learning to Classify Handwritten Digits**

* Using Deep Learning to Classify Handwritten Digits
* Understanding the McCulloch-Pitts neuron
* Understanding the perceptron
* Implementing your first perceptron
* Understanding multilayer perceptrons
* Getting acquainted with deep learning
* Classifying handwritten digits

1. **Combining Different Algorithms into an Ensemble**

* Combining Different Algorithms into an Ensemble
* Understanding ensemble methods
* Combining decision trees into a random forest
* Using random forests for face recognition
* Implementing AdaBoost
* Combining different models into a voting classifier

1. **Selecting the Right Model with Hyperparameter Tuning**

* Selecting the Right Model with Hyperparameter Tuning
* Evaluating a model
* Understanding cross-validation
* Estimating robustness using bootstrapping
* Assessing the significance of our results
* Tuning hyperparameters with grid search
* Scoring models using different evaluation metrics
* Chaining algorithms together to form a pipeline

1. **Wrapping Up**

* Wrapping Up
* Approaching a machine learning problem
* Building your own estimator
* Where to go from here?

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning for the Web

Course Snapshot

* **Course:** Machine Learning for the Web
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning for the Web skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Explore the web and make smarter predictions using Python
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Python is a general purpose and also a comparatively easy to learn programming language. Hence it is the language of choice for data scientists to prototype, visualize, and run data analyses on small and medium-sized data sets. This is a unique course that helps bridge the gap between machine learning and web development. It focuses on the difficulties of implementing predictive analytics in web applications. We focus on the Python language, frameworks, tools, and libraries, showing you how to build a machine learning system. You will explore the core machine learning concepts and then develop and deploy the data into a web application using the Django framework. You will also learn to carry out web, document, and server mining tasks, and build recommendation engines. Later, you will explore Python’s impressive Django framework and will find out how to build a modern simple web app with machine learning features.

Working in a hands-on learning environment, led by our Machine Learning for the Web instructor, students will learn about and explore:

* Targets two big and prominent markets where sophisticated web apps are of need and importance.
* Practical examples of building machine learning web application, which are easy to follow and replicate.
* A comprehensive tutorial on Python libraries and frameworks to get you up and started.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get familiar with the fundamental concepts and some of the jargons used in the machine learning community
* Use tools and techniques to mine data from websites
* Grasp the core concepts of Django framework
* Get to know the most useful clustering and classification techniques and implement them in Python
* Acquire all the necessary knowledge to build a web application with Django
* Successfully build and deploy a movie recommendation system application using the Django framework in Python

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to Explore the web and make smarter predictions using Python

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Introduction to Practical Machine Learning Using Python**

* Introduction to Practical Machine Learning Using Python
* General machine-learning concepts
* Preparing, manipulating and visualizing data – NumPy, pandas and matplotlib tutorials
* Scientific libraries
* When to use machine learning

1. **Unsupervised Machine Learning**

* Unsupervised Machine Learning
* Clustering algorithms
* Dimensionality reduction
* Singular value decomposition

1. **Supervised Machine Learning**

* Supervised Machine Learning
* Model error estimation
* Generalized linear models
* Naive Bayes
* Decision trees
* Support vector machine
* A comparison of methods
* Hidden Markov model

1. **Web Mining Techniques**

* Web Mining Techniques
* Web structure mining
* Web content mining
* Natural language processing
* Postprocessing information

1. **Recommendation Systems**

* Recommendation Systems
* Utility matrix
* Similarities measures
* Collaborative Filtering methods
* CBF methods
* Association rules for learning recommendation system
* Log-likelihood ratios recommendation system method
* Hybrid recommendation systems
* Evaluation of the recommendation systems

1. **Getting Started with Django**

* Getting Started with Django
* HTTP – the basics of the GET and POST methods
* Writing an app – most important features
* Admin

1. **Movie Recommendation System Web Application**

* Movie Recommendation System Web Application
* Application setup
* Models
* Commands
* User sign up login/logout implementation
* Information retrieval system (movies query)
* Rating system
* Recommendation systems
* Admin interface and API

1. **Sentiment Analyser Application for Movie Reviews**

* Sentiment Analyser Application for Movie Reviews
* Application usage overview
* Search engine choice and the application code
* Scrapy setup and the application code
* Django models
* Integrating Django with Scrapy
* PageRank: Django view and the algorithm code
* Admin and API

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Java Machine Learning

Course Snapshot

* **Course:** Java Machine Learning
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Become an advanced practitioner with this progressive set of master classes on application-oriented machine learning
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Java is one of the main languages used by practicing data scientists; much of the Hadoop ecosystem is Java-based, and it is certainly the language that most production systems in Data Science are written in. If you know Java, Mastering Machine Learning with Java is your next step on the path to becoming an advanced practitioner in Data Science. This course aims to introduce you to an array of advanced techniques in machine learning, including classification, clustering, anomaly detection, stream learning, active learning, semi-supervised learning, probabilistic graph modeling, text mining, deep learning, and big data batch and stream machine learning. Accompanying each lesson are illustrative examples and real-world case studies that show how to apply the newly learned techniques using sound methodologies and the best Java-based tools available today. On completing this course, you will understand the tools and techniques for building powerful machine learning models to solve data science problems in just about any domain.

Working in a hands-on learning environment, led by our Java Machine Learning instructor, students will learn about and explore:

* Comprehensive coverage of key topics in machine learning with an emphasis on both the theoretical and practical aspects
* More than 15 open source Java tools in a wide range of techniques, with code and practical usage.
* More than 10 real-world case studies in machine learning highlighting techniques ranging from data ingestion up to analyzing the results of experiments, all preparing the user for the practical, real-world use of tools and data analysis

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Master key Java machine learning libraries, and what kind of problem each can solve, with theory and practical guidance.
* Explore powerful techniques in each major category of machine learning such as classification, clustering, anomaly detection, graph modeling, and text mining.
* Apply machine learning to real-world data with methodologies, processes, applications, and analysis.
* Techniques and experiments developed around the latest specializations in machine learning, such as deep learning, stream data mining, and active and semi-supervised learning.
* Build high-performing, real-time, adaptive predictive models for batch- and stream-based big data learning using the latest tools and methodologies.
* Get a deeper understanding of technologies leading towards a more powerful AI applicable in various domains such as Security, Financial Crime, Internet of Things, social networking, and so on.

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to Become an advanced practitioner with this progressive set of master classes on application-oriented machine learning

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Machine Learning Review**

* Machine Learning Review
* Machine learning – history and definition
* What is not machine learning?
* Machine learning – concepts and terminology
* Machine learning – types and subtypes
* Datasets used in machine learning
* Machine learning applications
* Practical issues in machine learning
* Machine learning – roles and process
* Machine learning – tools and datasets

1. **Practical Approach to Real-World Supervised Learning**

* Practical Approach to Real-World Supervised Learning
* Formal description and notation
* Data transformation and preprocessing
* Feature relevance analysis and dimensionality reduction
* Model building
* Model assessment, evaluation, and comparisons
* Case Study – Horse Colic Classification

1. **Unsupervised Machine Learning Techniques**

* Unsupervised Machine Learning Techniques
* Issues in common with supervised learning
* Issues specific to unsupervised learning
* Feature analysis and dimensionality reduction
* Clustering
* Outlier or anomaly detection
* Real-world case study

1. **Semi-Supervised and Active Learning**

* Semi-Supervised and Active Learning
* Semi-supervised learning
* Active learning
* Case study in active learning

1. **Real-Time Stream Machine Learning**

* Real-Time Stream Machine Learning
* Assumptions and mathematical notations
* Basic stream processing and computational techniques
* Concept drift and drift detection
* Incremental supervised learning
* Incremental unsupervised learning using clustering
* Unsupervised learning using outlier detection
* Case study in stream learning

1. **Probabilistic Graph Modeling**

* Probabilistic Graph Modeling
* Probability revisited
* Graph concepts
* Bayesian networks
* Markov networks and conditional random fields
* Specialized networks
* Tools and usage

1. **Deep Learning**

* Deep Learning
* Multi-layer feed-forward neural network
* Limitations of neural networks
* Deep learning

1. **Text Mining and Natural Language Processing**

* Text Mining and Natural Language Processing
* NLP, subfields, and tasks
* Issues with mining unstructured data
* Text processing components and transformations
* Topics in text mining
* Tools and usage

1. **Big Data Machine Learning – The Final Frontier**

* Big Data Machine Learning – The Final Frontier
* What are the characteristics of Big Data?
* Big Data Machine Learning
* Batch Big Data Machine Learning
* Case study

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Predictive Analytics with R

Course Snapshot

* **Course:** Predictive Analytics with R
* **Duration:** 4 days
* **Skill-level**: Foundation-level R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to master the craft of predictive modeling in R by developing strategy, intuition, and a solid foundation in essential concepts
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

R offers a free and open source environment that is perfect for both learning and deploying predictive modeling solutions. With its constantly growing community and plethora of packages, R offers the functionality to deal with a truly vast array of problems.

The course begins with a dedicated chapter on the language of models and the predictive modeling process. You will understand the learning curve and the process of tidying data. Each subsequent chapter tackles a particular type of model, such as neural networks, and focuses on the three important questions of how the model works, how to use R to train it, and how to measure and assess its performance using real-world datasets. How do you train models that can handle really large datasets? This course will also show you just that. Finally, you will tackle the really important topic of deep learning by implementing applications on word embedding and recurrent neural networks. By the end of this course, you will have explored and tested the most popular modeling techniques in use on real- world datasets and mastered a diverse range of techniques in predictive analytics using R.

Working in a hands-on learning environment, led by our R instructor, students will learn about and explore:

* Grasping the major methods of predictive modeling and moving beyond black box thinking to a deeper level of understanding
* Leveraging the flexibility and modularity of R to experiment with a range of different techniques and data types
* Packed with practical advice and tips explaining important concepts and best practices to help you understand quickly and easily

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Master the steps involved in the predictive modeling process
* Grow your expertise in using R and its diverse range of packages
* Learn how to classify predictive models and distinguish which models are suitable for a particular problem
* Understand steps for tidying data and improving the performing metrics
* Recognize the assumptions, strengths, and weaknesses of a predictive model
* Understand how and why each predictive model works in R
* Select appropriate metrics to assess the performance of different types of predictive model
* Explore word embedding and recurrent neural networks in R
* Train models in R that can work on very large datasets

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to master the craft of predictive modeling in R by developing strategy, intuition, and a solid foundation in essential concepts

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Gearing Up for Predictive** Modeling

* Gearing Up for Predictive Modeling
* Models
* Types of model
* The process of predictive modeling

1. **Tidying Data and Measuring Performance**

* Tidying Data and Measuring Performance
* Getting started
* Tidying data
* Categorizing data quality
* Performance metrics
* Cross-validation
* Learning curves

1. **Linear Regression**

* Linear Regression
* Introduction to linear regression
* Simple linear regression
* Multiple linear regression
* Assessing linear regression models
* Problems with linear regression
* Feature selection
* Regularization
* Polynomial regression

1. **Generalized Linear Models**

* Generalized Linear Models
* Classifying with linear regression
* Introduction to logistic regression
* Predicting heart disease
* Assessing logistic regression models
* Regularization with the lasso
* Classification metrics
* Extensions of the binary logistic classifier
* Poisson regression
* Negative Binomial regression

1. **Neural Networks**

* Neural Networks
* The biological neuron
* The artificial neuron
* Stochastic gradient descent
* Multilayer perceptron networks
* The back propagation algorithm
* Predicting the energy efficiency of buildings
* Predicting glass type revisited
* Predicting handwritten digits
* Radial basis function networks

1. **Support Vector Machines**

* Support Vector Machines
* Maximal margin classification
* Support vector classification
* Kernels and support vector machines
* Predicting chemical biodegration
* Predicting credit scores
* Multiclass classification with support vector machines

1. **Tree-Based Methods**

* Tree-Based Methods
* The intuition for tree models
* Algorithms for training decision trees
* Predicting class membership on synthetic 2D data
* Predicting the authenticity of banknotes
* Predicting complex skill learning
* Improvements to the M5 model

1. **Dimensionality Reduction**

* Dimensionality Reduction
* Defining DR

1. **Ensemble Methods**

* Ensemble Methods
* Bagging
* Boosting
* Predicting atmospheric gamma ray radiation
* Predicting complex skill learning with boosting

1. **Probabilistic Graphical Models**

* Probabilistic Graphical Models
* A little graph theory
* Bayes' theorem
* Conditional independence
* Bayesian networks
* The Naïve Bayes classifier

1. **Topic Modeling**

* Topic Modeling
* An overview of topic modeling
* Latent Dirichlet Allocation
* Modeling the topics of online news stories
* Modeling tweet topics

1. **Recommendation Systems**

* Recommendation Systems
* Rating matrix
* Collaborative filtering
* Singular value decomposition
* Predicting recommendations for movies and jokes
* Loading and pre-processing the data
* Exploring the data
* Other approaches to recommendation systems

1. **Scaling Up**

* Scaling Up
* Starting the project
* Characteristics of big data
* Training models at scale
* A path forward
* Alternatives

1. **Deep Learning**

* Deep Learning
* Machine learning or deep learning
* What is deep learning?

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Statistical Application Development with R and Python

Course Snapshot

* **Course:** Statistical Application Development with R and Python
* **Duration:** 3 days
* **Skill-level**: Foundation-level R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending Software Implementation Illustrated with R and Python
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Statistical Analysis involves collecting and examining data to describe the nature of data that needs to be analyzed. It helps you explore the relation of data and build models to make better decisions. This course explores statistical concepts along with R and Python, which are well integrated from the word go. Almost every concept has an R code going with it which exemplifies the strength of R and applications. The R code and programs have been further strengthened with equivalent Python programs. Thus, you will first understand the data characteristics, descriptive statistics and the exploratory attitude, which will give you firm footing of data analysis. Statistical inference will complete the technical footing of statistical methods. Regression, linear, logistic modeling, and CART, builds the essential toolkit. This will help you complete complex problems in the real world. You will begin with a brief understanding of the nature of data and end with modern and advanced statistical models like CART. Every step is taken with DATA and R code, and further enhanced by Python. The data analysis journey begins with exploratory analysis, which is more than simple, descriptive, data summaries. You will then apply linear regression modeling, and end with logistic regression, CART, and spatial statistics. By the end of this course you will be able to apply your statistical learning in major domains at work or in your projects.

Working in a hands-on learning environment, led by our R and Python instructor, students will learn about and explore:

* Learn the nature of data through software which takes the preliminary concepts right away using R and Python.
* Understand data modeling and visualization to perform efficient statistical analysis with this guide.
* Get well versed with techniques such as regression, clustering, classification, support vector machines and much more to learn the fundamentals of modern statistics.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn the nature of data through software with preliminary concepts right away in R
* Read data from various sources and export the R output to other software
* Perform effective data visualization with the nature of variables and rich alternative options
* Do exploratory data analysis for useful first sight understanding building up to the right attitude towards effective inference
* Learn statistical inference through simulation combining the classical inference and modern computational power
* Delve deep into regression models such as linear and logistic for continuous and discrete regressands for forming the fundamentals of modern statistics
* Introduce yourself to CART – a machine learning tool which is very useful when the data has an intrinsic nonlinearity

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to do a Software Implementation Illustrated with R and Python

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Data Characteristics**

* Data Characteristics
* Questionnaire and its components
* Experiments with uncertainty in computer science
* Installing and setting up R
* Using R packages
* Python installation and setup
* IDEs for R and Python
* The companion code bundle
* Discrete distributions
* Continuous distributions

1. **Import/Export Data**

* Import/Export Data
* Packages and settings – R and Python
* Understanding data.frame and other formats
* Using utils and the foreign packages
* Exporting data/graphs

1. **Data Visualization**

* Data Visualization
* Packages and settings – R and Python
* Visualization techniques for categorical data
* Visualization techniques for continuous variable data
* Pareto chart
* A brief peek at ggplot2

1. **Exploratory Analysis**

* Exploratory Analysis
* Packages and settings – R and Python
* Essential summary statistics
* Techniques for exploratory analysis

1. **Statistical Inference**

* Statistical Inference
* Packages and settings – R and Python
* Maximum likelihood estimator
* Confidence intervals
* Hypothesis testing

1. **Linear Regression Analysis**

* Linear Regression Analysis
* Packages and settings - R and Python
* The essence of regression
* The simple linear regression model
* Multiple linear regression model
* Regression diagnostics
* Model selection

1. **Logistic Regression Model**

* Logistic Regression Model
* Packages and settings – R and Python
* Model validation and diagnostics
* Logistic regression for the German credit screening dataset

1. **Regression Models with Regularization**

* Regression Models with Regularization
* Packages and settings – R and Python
* Regression spline
* Ridge regression for linear models

1. **Classification and Regression Trees**

* Classification and Regression Trees
* Packages and settings – R and Python
* Splitting the data

1. **CART and Beyond**

* CART and Beyond
* Packages and settings – R and Python
* Understanding bagging

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Learn Unity ML-Agents

Course Snapshot

* **Course:** Learn Unity ML-Agents
* **Duration:** 2 days
* **Skill-level**: Foundation-level Learn Unity ML-Agents skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wish to Transform games into environments using machine learning and Deep learning with Tensorflow, Keras, and Unity
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Unity Machine Learning agents allow researchers and developers to create games and simulations using the Unity Editor, which serves as an environment where intelligent agents can be trained with machine learning methods through a simple-to-use Python API. This course takes you from the basics of Reinforcement and Q Learning to building Deep Recurrent Q-Network agents that cooperate or compete in a multi-agent ecosystem. You will start with the basics of Reinforcement Learning and how to apply it to problems. Then you will learn how to build self-learning advanced neural networks with Python and Keras/TensorFlow. From there you move o n to more advanced training scenarios where you will learn further innovative ways to train your network with A3C, imitation, and curriculum learning models. By the end of the course, you will have learned how to build more complex environments by building a cooperative and competitive multi-agent ecosystem.

Working in a hands-on learning environment, led by our Unity ML-Agents instructor, students will learn about and explore:

* Learn how to apply core machine learning concepts to your games with Unity
* Learn the Fundamentals of Reinforcement Learning and Q-Learning and apply them to your games
* Learn How to build multiple asynchronous agents and run them in a training scenario

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Develop Reinforcement and Deep Reinforcement Learning for games.
* Understand complex and advanced concepts of reinforcement learning and neural networks
* Explore various training strategies for cooperative and competitive agent development
* Adapt the basic script components of Academy, Agent, and Brain to be used with Q Learning.
* Enhance the Q Learning model with improved training strategies such as Greedy-Epsilon exploration
* Implement a simple NN with Keras and use it as an external brain in Unity
* Understand how to add LTSM blocks to an existing DQN
* Build multiple asynchronous agents and run them in a training scenario

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to Transform games into environments using machine learning and Deep learning with Tensorflow, Keras, and Unity

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Introducing Machine Learning and ML-Agents**

* Introducing Machine Learning and ML-Agents
* Machine Learning
* ML-Agents
* Running a sample
* Creating an environment
* Academy, Agent, and Brain

1. **The Bandit and Reinforcement Learning**

* The Bandit and Reinforcement Learning
* Reinforcement Learning
* Contextual bandits and state
* Exploration and exploitation
* MDP and the Bellman equation
* Q-Learning and connected agents

1. **Deep Reinforcement Learning with Python**

* Deep Reinforcement Learning with Python
* Installing Python and tools
* ML-Agents external brains
* Neural network foundations
* Deep Q-learning
* Proximal policy optimization

1. **Going Deeper with Deep Learning**

* Going Deeper with Deep Learning
* Agent training problems
* Convolutional neural networks
* Experience replay
* Partial observability, memory, and recurrent networks
* Asynchronous actor – critic training

1. **Playing the Game**

* Playing the Game
* Multi-agent environments
* Adversarial self-play
* Decisions and On-Demand Decision Making
* Imitation learning
* Curriculum Learning

1. **Terrarium Revisited – A Multi-Agent Ecosystem**

* Terrarium Revisited – A Multi-Agent Ecosystem
* What was/is Terrarium?
* Building the Agent ecosystem
* Basic Terrarium – Plants and Herbivores
* Carnivore: the hunter

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# TensorFlow 2.0

Course Snapshot

* **Course:** TensorFlow 2.0
* **Duration:** 2 days
* **Skill-level**: Foundation-level TensorFlow skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Perform supervised and unsupervised machine learning and learn advanced techniques such as training neural networks.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

TensorFlow is one of the most popular machine learning frameworks in Python. With this course, you will improve your knowledge of some of the latest TensorFlow features and will be able to perform supervised and unsupervised machine learning and also train neural networks. After giving you an overview of what's new in TensorFlow 2.0 Alpha, the course moves on to setting up your machine learning environment using the TensorFlow library. You will perform popular supervised machine learning tasks using techniques such as linear regression, logistic regression, and clustering. You will get familiar with unsupervised learning for autoencoder applications. The course will also show you how to train effective neural networks using straightforward examples in a variety of different domains. By the end of the course, you will have been exposed to a large variety of machine learning and neural network TensorFlow techniques.

Working in a hands-on learning environment, led by our TensorFlow 2.0 expert instructor, students will learn about and explore:

* Train your own models for effective prediction, using high-level Keras API
* Perform supervised and unsupervised machine learning and learn advanced techniques such as training neural networks
* Get acquainted with some new practices introduced in TensorFlow 2.0 Alpha

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use tf.Keras for fast prototyping, building, and training deep learning neural network models
* Easily convert your TensorFlow 1.12 applications to TensorFlow 2.0-compatible files
* Use TensorFlow to tackle traditional supervised and unsupervised machine learning applications
* Understand image recognition techniques using TensorFlow
* Perform neural style transfer for image hybridization using a neural network
* Code a recurrent neural network in TensorFlow to perform text-style generation

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to Perform supervised and unsupervised machine learning and learn advanced techniques such as training neural networks.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Introducing TensorFlow 2**

* Introducing TensorFlow 2
* Looking at the modern TensorFlow ecosystem
* Installing TensorFlow
* Housekeeping and eager operations
* Providing useful TensorFlow operations

1. **Keras, a High-Level API for TensorFlow 2**

* Keras, a High-Level API for TensorFlow 2
* The adoption and advantages of Keras
* The features of Keras
* The default Keras configuration file
* The Keras backend
* Keras data types
* Keras models

1. **ANN Technologies Using TensorFlow 2**

* ANN Technologies Using TensorFlow 2
* Presenting data to an ANN
* One-hot encoding
* Layers
* Activation functions
* Creating the model
* Gradient calculations for gradient descent algorithms
* Loss functions

1. **Supervised Machine Learning Using TensorFlow 2**

* Supervised Machine Learning Using TensorFlow 2
* Supervised learning
* Linear regression
* Our first linear regression example
* The Boston housing dataset
* Logistic regression (classification)
* k-Nearest Neighbors (KNN)

1. **Unsupervised Learning Using TensorFlow 2**

* Unsupervised Learning Using TensorFlow 2
* Autoencoders

1. **Recognizing Images with TensorFlow 2**

* Recognizing Images with TensorFlow 2
* Quick Draw – image classification using TensorFlow
* CIFAR 10 image classification using TensorFlow

1. **Neural Style Transfer Using TensorFlow 2**

* Neural Style Transfer Using TensorFlow 2
* Setting up the imports
* Preprocessing the images
* Viewing the original images
* Using the VGG19 architecture
* Creating the model
* Calculating the losses
* Performing the style transfer
* Final displays

1. **Recurrent Neural Networks Using TensorFlow 2**

* Recurrent Neural Networks Using TensorFlow 2
* Neural network processing modes
* Recurrent architectures
* An application of RNNs
* The code for our RNN example
* Building and instantiating our model
* Using our model to get predictions

1. **TensorFlow Estimators and TensorFlow Hub**

* TensorFlow Estimators and TensorFlow Hub
* TensorFlow Estimators
* TensorFlow Hub

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Deep Learning for Vision Systems

Course Snapshot

* **Course:** Deep Learning for Vision Systems
* **Duration:** 2 days
* **Skill-level**: Foundation-level Deep Learning for Vision Systems skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python developers, analysts or others who wants to get concepts and tools for building intelligent, scalable computer vision systems that can identify and react to objects in images, videos, and real life
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Deep Learning for Vision Systems** teaches you to apply deep learning techniques to solve real-world computer vision problems. In his straightforward and accessible style, DL and CV expert Mohamed Elgendy introduces you to the concept of visual intuition—how a machine learns to understand what it sees. Then you’ll explore the DL algorithms used in different CV applications. You’ll drill down into the different parts of the CV interpreting system, or pipeline. Using Python, OpenCV, Keras, TensorFlow, and Amazon’s Mx Net, you’ll discover advanced DL techniques for solving CV problems. Applications of focus include image classification, segmentation, captioning, and generation as well as face recognition and analysis. You’ll also cover the most important deep learning architectures including artificial neural networks (ANNs), convolutional networks (CNNs), and recurrent networks (RNNs), knowledge that you can apply to related deep learning disciplines like natural language processing and voice user interface. Real-life, scalable projects from Amazon, Google, and Facebook drive it all home. With this invaluable course, you’ll gain the essential skills for building amazing end-to-end CV projects that solve real-world problems.

Working in a hands-on learning environment, led by our Deep Learning expert instructor, students will learn about and explore:

* explore the DL algorithms used in different CV applications.
* You’ll drill down into the different parts of the CV interpreting system, or pipeline. Using Python, OpenCV, Keras, TensorFlow, and Amazon’s Mx Net,
* you’ll discover advanced DL techniques for solving CV problems.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Introduction to computer vision
* Deep learning and neural networks
* Transfer learning and advanced CNN architectures
* Image classification and captioning
* Object detection with YOLO, SSD and R-CNN
* Style transfer
* AI ethics
* Real-world projects

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to get concepts and tools for building intelligent, scalable computer vision systems that can identify and react to objects in images, videos, and real life

**Pre-Requisites:** Students should have

* with intermediate Python, math and machine learning skills.
* Experience with the Matplotlib and Pandas machine learning libraries is helpful
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Welcome to Computer Vision**

* Computer vision intuition
* Applications of computer vision
* Computer Vision Pipeline - The big picture
* Input image
* Image preprocessing
* Feature extraction
* Classifier learning algorithm
* Ch summary and takeaways

1. **Deep learning and neural networks**

* The Perceptron intuition
* Multi-Layer Perceptron (MLP)
* Activation functions
* Feedforward
* Error functions
* Optimization algorithms
* Backpropagation
* Ch summary and takeaways
* Project: Build Your first Neural Network

1. **Convolutional Neural Networks (CNNs)**

* Image classification using MLP
* CNNs Architecture
* Basic components of the CNN
* Image classification using CNNs
* Add Dropout layers to avoid overfitting
* Convolution over colored images (3D images)
* Ch summary and takeaways
* Project: Image classification for colored images (CIFAR-10 dataset)

1. **Structuring Deep Learning Projects and Hyperparameters tuning**

* Define the performance metrics
* Design a baseline model
* Get your data ready for training
* Evaluate the model and interpret its performance (error analysis)
* Improve the network and tune hyperparameters
* Batch normalization (BN)
* Ch summary and takeaways
* Project: Achieve >90% accuracy on the CIFAR-10 image classification project

1. **Advanced CNN Architectures**

* CNN design patterns
* LeNet-5
* VGGNet
* Inception and Google Net
* Res Net

1. **Transfer Learning**

* What are the problems that transfer learning is solving?
* What is transfer learning?
* How transfer learning works
* Transfer learning approaches
* Choose the appropriate level of transfer learning
* Open-source datasets
* Ch summary and takeaways
* Project 1: A pretrained network as a feature extractor
* Project 2: Fine tuning

1. **Object Detection with R-CNN, SSD, and YOLO**

* General object detection framework
* Region-Based Convolutional Neural Networks (R-CNNs)
* Single Shot Detection (SSD)
* Ch summary and takeaways

1. **Generative Adversarial Networks (GANs)**

* GANs Architecture
* Evaluate GAN models
* Popular GANs Applications
* Building your own GAN project

1. **Deep Dream and Neural Style Transfer**

* How convolutional neural networks see the world
* Deep Dream
* Neural Style Transfer

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Deep Learning with PyTorch

Course Snapshot

* **Course:** Deep Learning with PyTorch
* **Duration:** 4 days
* **Skill-level**: Foundation-level Deep Learning with PyTorch skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are build and train the latest and greatest deep learning models and contribute to making a dent in the world.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Deep Learning with PyTorch teaches you how to implement deep learning algorithms with Python and PyTorch. This course takes you into a fascinating case study: building an algorithm capable of detecting malignant lung tumors using CT scans. As the authors guide you through this real example, you'll discover just how effective and fun PyTorch can be. After a quick introduction to the deep learning landscape, you'll explore the use of pre-trained networks and start sharpening your skills on working with tensors. You'll find out how to represent the most common types of data with tensors and how to build and train neural networks from scratch on practical examples, focusing on images and sequences. After covering the basics, the course will take you on a journey through larger projects. The centerpiece of the course is a neural network designed for cancer detection. You'll discover ways for training networks with limited inputs and start processing data to get some results. You'll sift through the unreliable initial results and focus on how to diagnose and fix the problems in your neural network. Finally, you'll look at ways to improve your results by training with augmented data, make improvements to the model architecture, and perform other fine tuning.!

Working in a hands-on learning environment, led by our Deep Learning with PyTorch expert instructor, students will learn about and explore:

* you'll explore the use of pre-trained networks and start sharpening your skills on working with tensors.
* You'll find out how to represent the most common types of data with tensors and how to build and train neural networks from scratch on practical examples, focusing on images and sequences.
* After covering the basics, the course will take you on a journey through larger projects.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Using the PyTorch tensor API
* Understanding automatic differentiation in PyTorch
* Training deep neural networks
* Monitoring training and visualizing results
* Implementing modules and loss functions
* Loading data in Python for PyTorch
* Interoperability with NumPy
* Deploying a PyTorch model for inference

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish that how to implement deep learning algorithms with Python and PyTorch

**Pre-Requisites:** Students should have

* developers with some knowledge of Python as well as basic linear algebra skills.
* Some understanding of deep learning will be helpful
* No experience with PyTorch or other deep learning frameworks is required.
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Introducing Deep Learning and the PyTorch Library**

* What is PyTorch?
* What is this course?
* Why PyTorch
* competitive landscape
* 1.4 PyTorch has the batteries included

1. **Pre-Trained Networks**

* A pre-trained network that recognizes the subject of an image
* A pre-trained model that fakes it until it makes it
* A pre-trained network that describes scenes
* Torch Hub

1. **Ch 3: It Starts with a Tensor**

* **Tensors are multi-dimensional arrays**
* Indexing Tensors
* Named Tensors
* Tensor element types
* The tensor API
* Tensors — scenic views on storage
* Tensor metadata: size, offset, stride
* NumPy interoperability
* Moving tensors to the GPU
* Generalized Tensors are Tensors, too
* Serializing tensors

1. **Real-World Data Representation Using Tensors**

* Images
* Volumetric Data
* Tabular Data
* Time Series
* Text

1. **The Mechanics of Learning**

* Learning is just parameter estimation
* PyTorch’s Auto grad: Back-propagate all things

1. **Using A Neural Network to Fit the Data**

* Artificial Neurons
* The PyTorch nn module
* Sub classing nn. Module

1. **Telling Birds from Airplanes: Learning from Images**

* A dataset of tiny images
* Distinguishing birds from airplanes

1. **Using Convolutions to Generalize**

* The case for convolutions
* Convolutions
* Sub classing nn. Module
* Training our Convnet
* Model Design

**Part 2: Learning from Images in the Real-World: Early Detection of Lung Cancer**

1. **Using PyTorch To Fight Cancer**

* What is a CT scan, exactly?
* The project: an end-to-end malignancy detector for lung cancer

1. **Ready, Dataset, Go!**

* Parsing LUNA’s annotation data
* Loading individual CT scans
* Locating a nodule using the patient coordinate system
* A straightforward Dataset implementation

1. **Training A Classification Model to Detect Suspected Tumors**

* The main entry point for our application
* Pre-training setup and initialization
* Our first-pass neural network design
* Training and validating the model
* Outputting performance metrics
* Running the training script
* Evaluating the model: Getting 99.7% correct means we’re done, right?
* Graphing training metrics with Tensor Board
* Why is the model not learning to detect malignant tumors?

1. **Monitoring Metrics: Precision, Recall, and Pretty Pictures**

* Good dogs versus bad guys: false positives and false negatives
* Graphing the positives and negatives
* What does an ideal data set look like?
* Revisiting the problem of over-fitting
* Data Augmentation

1. **Using Segmentation to Find Suspected Nodules**

* Segmentation is per-pixel classification
* A 3D Dataset in 2D
* Updating the training script

1. **Deploying to production**

* Serving PyTorch models
* Exporting Models
* Interacting with the PyTorch JIT
* LibTorch — PyTorch in C++
* Emerging Technology: Enterprise serving of PyTorch models
* Going mobile

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Training Systems using Python Statistical Modeling

Course Snapshot

* **Course:** Training Systems using Python Statistical Modeling
* **Duration:** 2 days
* **Skill-level**: Foundation-level Training Systems using Python Statistical Modeling skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Leverage the power of Python and statistical modeling techniques for building accurate predictive models.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Python's ease of use and multi-purpose nature has led it to become the choice of tool for many data scientists and machine learning developers today. Its rich libraries are widely used for data analysis, and more importantly, for building state-of-the-art predictive models. This course takes you through an exciting journey, of using these libraries to implement effective statistical models for predictive analytics. You’ll start by diving into classical statistical analysis, where you will learn to compute descriptive statistics using pandas. You will look at supervised learning, where you will explore the principles of machine learning and train different machine learning models from scratch. You will also work with binary prediction models, such as data classification using k-nearest neighbors, decision trees, and random forests. This course also covers algorithms for regression analysis, such as ridge and lasso regression, and their implementation in Python. You will also learn how neural networks can be trained and deployed for more accurate predictions, and which Python libraries can be used to implement them. By the end of this course, you will have all the knowledge you need to design, build, and deploy enterprise-grade statistical models for machine learning using Python and its rich ecosystem of libraries for predictive analytics.

Working in a hands-on learning environment, led by our Deep Learning with PyTorch expert instructor, students will learn about and explore:

* Get introduced to Python's rich suite of libraries for statistical modeling
* Implement regression, clustering and train neural networks from scratch
* Includes real-world examples on training end-to-end machine learning systems in Python.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the importance of statistical modeling
* Learn about the various Python packages for statistical analysis
* Implement algorithms such as Naive Bayes, random forests, and more
* Build predictive models from scratch using Python's scikit-learn library
* Implement regression analysis and clustering
* Learn how to train a neural network in Python

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to Leverage the power of Python and statistical modeling techniques for building accurate predictive models

**Pre-Requisites:** Students should have

* developers with some knowledge of Python as well as basic linear algebra skills.
* Some understanding of deep learning will be helpful
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Classical Statistical Analysis**

* Classical Statistical Analysis
* Technical requirements
* Computing descriptive statistics
* Classical inference for proportions
* Classical inference for means
* Diving into Bayesian analysis
* Bayesian analysis for proportions
* Bayesian analysis for means
* Finding correlations

1. **Introduction to Supervised Learning**

* Introduction to Supervised Learning
* Principles of machine learning
* Training models
* Evaluating models

1. **Binary Prediction Models**

* Binary Prediction Models
* K-nearest neighbors classifier
* Decision trees
* Random forests
* Naive Bayes classifier
* Support vector machines
* Logistic regression
* Extending beyond binary classifiers

1. **Regression Analysis and How to Use It**

* Regression Analysis and How to Use It
* Linear models
* Evaluating linear models
* Bayesian linear models
* Ridge regression
* LASSO regression
* Spline interpolation

1. **Neural Networks**

* Neural Networks
* An introduction to perceptrons
* Neural networks
* MLPs for classification
* MLP for regression

1. **Clustering Techniques**

* Clustering Techniques
* Introduction to clustering
* Exploring the k-means algorithm
* Evaluating clusters
* Hierarchical clustering
* Spectral clustering

1. **Dimensionality Reduction**

* Dimensionality Reduction
* Introducing dimensionality reduction
* Principal component analysis
* Singular value decomposition
* Low-dimensional representation

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Deep Reinforcement Learning

**Course Snapshot**

* **Course:** Deep Reinforcement Learning
* **Duration:** 3 days
* **Skill-level**: Foundation-level Deep Reinforcement Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python developers, analysts or others who wants a powerful machine learning approach, using examples, illustrations, exercises, and crystal-clear teaching. You'll love the perfectly paced teaching and the clever, engaging writing style as you dig into this awesome exploration of reinforcement learning fundamentals, effective deep learning techniques, and practical applications in this emerging field.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Deep Reinforcement Learning** is a beautifully balanced approach to teaching, offering numerous large and small examples, annotated diagrams and code, engaging exercises, and skillfully crafted writing. You'll explore, discover, and learn as you lock in the ins and outs of reinforcement learning, neural networks, and AI agents. You will go from small grid world environments and some of the foundational algorithms to some of the most challenging environments out there today and cutting-edge techniques to solve these environments.

Working in a hands-on learning environment, led by our Deep Reinforcement expert instructor, students will learn about and explore:

* You'll explore, discover, and learn as you lock in the ins and outs of reinforcement learning, neural networks, and AI agents.
* You will go from small grid world environments and some of the foundational algorithms to some of the most challenging environments out there today and cutting-edge techniques to solve these environments.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Foundational reinforcement learning concepts and methods
* The most popular deep reinforcement learning agents solving high-dimensional environments
* Cutting-edge agents that emulate human-like behavior and techniques for artificial general intelligence

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to build machine learning systems that explore and learn based on the responses of the environment.

**Pre-Requisites:** Students should have

* developers with some understanding of deep learning algorithms.
* Experience with reinforcement learning is not required.
* Perfect for readers of Deep Learning in Python or Deep Learning.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction to deep reinforcement learning**

* What is deep reinforcement learning?
* The past, present, and future of deep reinforcement learning
* The suitability of deep reinforcement learning
* Setting clear two-way expectations

1. **Mathematical foundations of reinforcement earning**

* Components of reinforcement learning
* MDPs: The engine of the environment

1. **Balancing immediate and long-term goals**

* The objective of a decision-making agent
* Planning optimal sequences of actions

1. **Balancing the gathering and utilization of information**

* The challenge of interpreting evaluative feedback
* Strategic exploration

1. **Evaluating agents' behaviors**

* Learning to estimate the value of policies
* Learning to estimate from multiple steps

1. **Improving agents' behaviors**

* The anatomy of reinforcement learning agents
* Learning to improve policies of behavior
* Decoupling behavior from learning

1. **Achieving goals more effectively and efficiently**

* Learning to improve policies using robust targets
* Agents that interact, learn and plan

1. **Introduction to value-based deep reinforcement learning**

* The kind of feedback deep reinforcement learning agents use
* Introduction to function approximation for reinforcement learning
* NFQ: The first attempt to value-based deep reinforcement learning

1. **More stable value-based methods**

* DQN: Making reinforcement learning more like supervised learning
* Using experience replay
* Double DQN: Mitigating the overestimation of action-value functions

1. **Sample-efficient value-based methods**

* Dueling DDQN: A reinforcement-learning-aware neural network architecture
* network brings to the table?
* PER: Prioritizing the replay of meaningful experiences

1. **Policy-gradient and actor-critic methods**

* REINFORCE: Outcome-based policy learning
* VPG: Learning a value function
* A3C: Parallel policy updates
* GAE: Robust advantage estimation
* A2C: Synchronous policy updates

1. **Advanced actor-critic methods**

* DDPG: Approximating a deterministic policy
* TD3: State-of-the-art improvements over DDPG
* SAC: Maximizing the expected return and entropy

1. **Towards artificial general intelligence**

* What was covered, and what notably wasn’t?
* More advanced concepts towards AGI
* Fairness, and Ethical Standards
* What happens next?

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Applied Unsupervised Learning with R

**Course Snapshot**

* **Course:** Applied Unsupervised Learning with R
* **Duration:** 2 days
* **Skill-level**: Foundation-level Applied Unsupervised Learning with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python developers, analysts or others who wants to Design clever algorithms that discover hidden patterns and draw responses from unstructured, unlabeled data
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Starting with the basics, Applied Unsupervised Learning with R explains clustering methods, distribution analysis, data encoders, and features of R that enable you to understand your data better and get answers to your most pressing business questions. This course begins with the most important and commonly used method for unsupervised learning - clustering - and explains the three main clustering algorithms - k-means, divisive, and agglomerative. Following this, you'll study market basket analysis, kernel density estimation, principal component analysis, and anomaly detection. You'll be introduced to these methods using code written in R, with further instructions on how to work with, edit, and improve R code. To help you gain a practical understanding, the course also features useful tips on applying these methods to real business problems, including market segmentation and fraud detection. By working through interesting activities, you'll explore data encoders and latent variable models. By the end of this course, you will have a better understanding of different anomaly detection methods, such as outlier detection, Mahalanobis distances, and contextual and collective anomaly detection.

Working in a hands-on learning environment, led by our Applied Unsupervised Learning with R expert instructor, students will learn about and explore:

* Build state-of-the-art algorithms that can solve your business' problems
* Learn how to find hidden patterns in your data
* Revise key concepts with hands-on exercises using real-world datasets.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Implement clustering methods such as k-means, agglomerative, and divisive
* Write code in R to analyze market segmentation and consumer behavior
* Estimate distribution and probabilities of different outcomes
* Implement dimension reduction using principal component analysis
* Apply anomaly detection methods to identify fraud
* Design algorithms with R and learn how to edit or improve code

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Design clever algorithms that discover hidden patterns and draw responses from unstructured, unlabeled data.

**Pre-Requisites:** Students should have

* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction to Clustering Methods**

* Introduction to Clustering Methods
* Introduction
* Introduction to Clustering
* Introduction to the Iris Dataset
* Introduction to k-means Clustering
* Introduction to k-means Clustering with Built-In Functions
* Introduction to Market Segmentation
* Introduction to k-medoids Clustering

1. **Advanced Clustering Methods**

* Advanced Clustering Methods
* Introduction
* Introduction to k-modes Clustering
* Introduction to Density-Based Clustering (DBSCAN)

1. **Probability Distributions**

* Probability Distributions
* Introduction
* Basic Terminology of Probability Distributions
* Introduction to Kernel Density Estimation
* Introduction to the Kolmogorov-Smirnov Test

1. **Dimension Reduction**

* Dimension Reduction
* Introduction
* Market Basket Analysis

1. **Data Comparison Methods**

* Data Comparison Methods
* Introduction
* Analytic Signatures
* Comparison of Signatures
* Latent Variable Models – Factor Analysis

1. **Anomaly Detection**

* Anomaly Detection
* Introduction
* Univariate Outlier Detection
* Kernel Density

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Essential Natural Language Processing

**Course Snapshot**

* **Course:** Essential Natural Language Processing
* **Duration:** 3 days
* **Skill-level**: Foundation-level Natural Language Processing skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for basic Python developers, analysts or others who want to upgrade your applications with functions and features like information extraction, user profiling, and automatic topic labeling, this is the course for you.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Essential Natural Language Processing** is a hands-on guide to NLP with practical techniques you can put into action right away. By following the numerous Python-based examples and real-world case studies, you’ll apply NLP to search applications, extracting meaning from text, sentiment analysis, user profiling, and more. When you’re done, you’ll have a solid grounding in NLP that will serve as a foundation for further learning.

Working in a hands-on learning environment, led by our Natural Language Processing expert instructor, students will learn about and explore:

* you’ll apply NLP to search applications, extracting meaning from text, sentiment analysis, user profiling, and more.
* you’ll have a solid grounding in NLP that will serve as a foundation for further learning.
* provides a concrete example with practical techniques that you can put into practice right away.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Extracting information from raw text
* Named entity recognition
* Automating summarization of key facts
* Topic labeling

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who want to upgrade your applications with functions and features like information extraction, user profiling, and automatic topic labeling.

**Pre-Requisites:** Students should have

* For beginners to NLP with basic Python skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction**

* A brief history of NLP
* Typical tasks

1. **Your first NLP example**

* Introducing NLP in practice:spam filtering
* Understanding the task
* Implementing your own spam filter
* Deploying your spam filter in practice

1. **Introduction to Information Search**

* Understanding the task
* Processing the data further
* Information weighing
* Practical use of the search algorithm

1. **Information Extraction**

* Use cases
* Understanding the task
* Detecting word types with part-of-speech tagging
* Understanding sentence structure with syntactic parsing
* Building your own Information Extraction algorithm

1. **Author Profiling as a Machine Learning Task**

* Understanding the task
* Machine Learning pipeline at a first glance
* A closer look at the machine learning pipeline

1. **Linguistic Feature Engineering for Author Profiling**

* Another close look at the machine learning pipeline
* Feature engineering for authorship attribution
* Practical use of authorship attribution and user profiling

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# GANs

**Course Snapshot**

* **Course:** GANs
* **Duration:** 3 days
* **Skill-level**: Foundation-level GANs for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for to produce photo-realistic faces and other media objects. With the potential to produce stunningly realistic animations or shocking deep fakes, GANs are a huge step forward in deep learning systems.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**GANs** teaches you to build and train your own Generative Adversarial Networks. You’ll start by creating simple generator and discriminator networks that are the foundation of GAN architecture. Then, following numerous hands-on examples, you’ll train GANs to generate high-resolution images, image-to-image translation, and targeted data generation. Along the way, you’ll find pro tips for making your system smart, effective, and fast.

Working in a hands-on learning environment, led by our GAN expert instructor, students will learn about and explore:

* you will explore incredible AI technology capable of creating images, sound, and videos that are indistinguishable from the “real thing.” By pitting two neural networks against each other—one to generate fakes and one to spot them—GANs rapidly learn to produce photo-realistic faces and other media objects.
* With the potential to produce stunningly realistic animations or shocking deep fakes, GANs are a huge step forward in deep learning systems.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Building your first GAN
* Handling the progressive growing of GANs
* Practical applications of GANs
* Troubleshooting your system

**Audience & Pre-Requisites**

This course is geared for those who wants to create simple generator and discriminator networks that are the foundation of GAN architecture

**Pre-Requisites:** Students should have

* For data professionals with intermediate Python skills
* The basics of deep learning–based image processing.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction to GANs free**

* What are Generative Adversarial Networks?
* How do GANs work?
* GANs
* Why study GANs?

1. **Intro to generative modeling with autoencoders**

* Introduction to generative modeling
* How do autoencoders function on a high level?
* What are autoencoders to GANs?
* What is an autoencoder made of?
* Usage of autoencoders
* Unsupervised learning
* Code is life
* Why did we try a GAN?

1. **Your first GAN: Generating handwritten digits**

* Foundations of GANs: Adversarial training
* The Generator and the Discriminator
* GAN training algorithm
* Tutorial: Generating handwritten digits

1. **Deep Convolutional GAN**

* Convolutional neural networks
* Brief history of the DCGAN
* Batch normalization
* normalization
* 4.4. Tutorial: Generating handwritten digits with DCGAN

1. **Training and common challenges: Ganing for success**

* Evaluation
* Training challenges
* Summary of game setups
* Training hacks

1. **Progressing with GANs**

* Latent space interpolation
* They grow up so fast
* normalization in the generator
* Summary of key innovations
* TensorFlow Hub and hands-on
* Practical applications

1. **Semi-Supervised GAN**

* Introducing the Semi-Supervised GAN
* Tutorial: Implementing a Semi-Supervised GAN
* Comparison to a fully supervised classifier

1. **Conditional GAN**

* Motivation
* What is Conditional GAN?
* Tutorial: Implementing a Conditional GAN

1. **Cycle GAN**

* Image-to-image translation
* Cycle-consistency loss: There and back aGAN
* Adversarial loss
* Identity loss
* Architecture
* Object-oriented design of GANs
* Tutorial: Cycle GAN
* Expansions, augmentations, and applications

1. **Adversarial examples**

* Context of adversarial examples
* Lies, damned lies, and distributions
* Use and abuse of training
* Signal and the noise
* Not all hope is lost
* Adversaries to GANs

1. **Practical applications of GANs**

* GANs in medicine
* GANs in fashion

1. **Looking ahead**

* Ethics
* GAN innovations
* Further reading
* Looking back and closing thoughts

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with Apache Spark

**Course Snapshot**

* **Course:** Machine Learning with Apache
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning with Apache Spark skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to combine advanced analytics including Machine Learning, Deep Learning Neural Networks and Natural Language Processing with modern scalable technologies including Apache Spark to derive actionable insights from Big Data in real-time
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Every person and every organization in the world manage data, whether they realize it or not. Data is used to describe the world around us and can be used for almost any purpose, from analyzing consumer habits to fighting disease and serious organized crime. Ultimately, we manage data in order to derive value from it, and many organizations around the world have traditionally invested in technology to help process their data faster and more efficiently. But we now live in an interconnected world driven by mass data creation and consumption where data is no longer rows and columns restricted to a spreadsheet, but an organic and evolving asset in its own right. With this realization comes major challenges for organizations: how do we manage the sheer size of data being created every second (think not only spreadsheets and databases, but also social media posts, images, videos, music, blogs and so on)? And once we can manage all of this data, how do we derive real value from it? The focus of Machine Learning with Apache Spark is to help us answer these questions in a hands-on manner. We introduce the latest scalable technologies to help us manage and process big data. We then introduce advanced analytical algorithms applied to real-world use cases in order to uncover patterns, derive actionable insights, and learn from this big data.

Working in a hands-on learning environment, led by our Machine Learning with Apache expert instructor, students will learn about and explore:

* Make a hands-on start in the fields of Big Data, Distributed Technologies and Machine Learning
* Learn how to design, develop and interpret the results of common Machine Learning algorithms
* Uncover hidden patterns in your data in order to derive real actionable insights and business value

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand how Spark fits in the context of the big data ecosystem
* Understand how to deploy and configure a local development environment using Apache Spark
* Understand how to design supervised and unsupervised learning models
* Build models to perform NLP, deep learning, and cognitive services using Spark ML libraries
* Design real-time machine learning pipelines in Apache Spark
* Become familiar with advanced techniques for processing a large volume of data by applying machine learning algorithms

**Audience & Pre-Requisites**

This course is geared for attendees with Apache knowledge who wish to know the latest scalable technologies to help us manage and process big data. We then introduce advanced analytical algorithms applied to real-world use cases in order to uncover patterns, derive actionable insights, and learn from this big data.

**Pre-Requisites:** Students should have

* Basic to big-data-and-business-intelligence Skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **The Big Data Ecosystem**

* A brief history of data
* Big data ecosystem

1. **Setting Up a Local Development Environment**

* CentOS Linux 7 virtual machine

1. **Artificial Intelligence and Machine Learning**

* Artificial intelligence
* Machine learning
* Deep learning
* NLP
* Cognitive computing
* Machine learning pipelines in Apache Spark

1. **Supervised Learning Using Apache Spark**

* Linear regression
* Logistic regression
* Classification and Regression Trees

1. **Unsupervised Learning Using Apache Spark**

* Clustering
* Principal component analysis

1. **Natural Language Processing Using Apache Spark**

* Feature transformers
* Feature extractors
* Case study – sentiment analysis

1. **Deep Learning Using Apache Spark**

* Artificial neural networks

1. **Real-Time Machine Learning Using Apache Spark**

* Distributed streamingplatform
* Distributed stream processing engines
* Stream processing pipeline

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with Core ML

**Course Snapshot**

* **Course:** Machine Learning with Apache
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Leverage the power of Apple's Core ML to create smart iOS apps
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Core ML is a popular framework by Apple, with APIs designed to support various machine learning tasks. It allows you to train your machine learning models and then integrate them into your iOS apps. Machine Learning with Core ML is a fun and practical guide that not only demystifies Core ML but also sheds light on machine learning. In this course, you’ll walk through realistic and interesting examples of machine learning in the context of mobile platforms (specifically iOS). You’ll learn to implement Core ML for visual-based applications using the principles of transfer learning and neural networks. Having got to grips with the basics, you’ll discover a series of seven examples, each providing a new use-case that uncovers how machine learning can be applied along with the related concepts. By the end of the course, you will have the skills required to put machine learning to work in their own applications, using the Core ML APIs

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Explore the concepts of machine learning and Apple’s Core ML APIs
* Use Core ML to understand and transform images and videos
* Exploit the power of using CNN and RNN in iOS applications

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand components of an ML project using algorithms, problems, and data
* Master Core ML by obtaining and importing machine learning model, and generate classes
* Prepare data for machine learning model and interpret results for optimized solutions
* Create and optimize custom layers for unsupported layers
* Apply CoreML to image and video data using CNN
* Learn the qualities of RNN to recognize sketches, and augment drawing
* Use Core ML transfer learning to execute style transfer on images

**Audience & Pre-Requisites**

This course is geared for attendees with Apache knowledge who wish to Leverage the power of Apple's Core ML to create smart iOS apps.

**Pre-Requisites:** Students should have

* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction to Machine Learning**

* Introduction to Machine Learning
* What is machine learning?
* A brief tour of ML algorithms
* A typical ML workflow

1. **Introduction to Apple Core ML**

* Introduction to Apple Core ML
* Difference between training and inference
* Inference on the edge
* A brief introduction to Core ML
* Learning algorithms
* Considerations

1. **Recognizing Objects in the World**

* Recognizing Objects in the World
* Understanding images
* Recognizing objects in the world
* Performing inference

1. **Emotion Detection with CNNs**

* Emotion Detection with CNNs
* Facial expressions
* Input data and preprocessing
* Bringing it all together

1. **Locating Objects in the World**

* Locating Objects in the World
* Object localization and object detection
* Converting Keras Tiny YOLO to Core ML
* Making it easier to find photos
* Optimizing with batches

1. **Creating Art with Style Transfer**

* Creating Art with Style Transfer
* Transferring style from one image to another
* A faster way to transfer style
* Converting a Keras model to Core ML
* Building custom layers in Swift
* Reducing your model's weight

1. **Assisted Drawing with CNNs**

* Assisted Drawing with CNNs
* Towards intelligent interfaces
* Drawing
* Recognizing the user's sketch

1. **Assisted Drawing with RNNs**

* Assisted Drawing with RNNs
* Assisted drawing
* Recurrent Neural Networks for drawing classification
* Input data and preprocessing
* Bringing it all together

1. **Object Segmentation Using CNNs**

* Object Segmentation Using CNNs
* Classifying pixels
* Data to drive the desired effect – action shots
* Building the photo effects application
* Working with probabilistic results

1. **An Introduction to Create ML**

* An Introduction to Create ML
* A typical workflow
* Preparing the data
* Creating and training a model
* Closing thoughts

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Python Machine Learning By Example

**Course Snapshot**

* **Course:** Python Machine Learning
* **Duration:** 3 days
* **Skill-level**: Foundation-level Python Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Grasp machine learning concepts, techniques, and algorithms with the help of real-world examples using Python libraries such as TensorFlow and scikit-learn
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The surge in interest in machine learning (ML) is due to the fact that it revolutionizes automation by learning patterns in data and using them to make predictions and decisions. If you’re interested in ML, this course will serve as your entry point to ML. Python Machine Learning By Example begins with an introduction to important ML concepts and implementations using Python libraries. Each lesson of the course walks you through an industry adopted application. You’ll implement ML techniques in areas such as exploratory data analysis, feature engineering, and natural language processing (NLP) in a clear and easy-to-follow way. With the help of this extended and updated edition, you’ll understand how to tackle data-driven problems and implement your solutions with the powerful yet simple Python language and popular Python packages and tools such as TensorFlow, scikit-learn, gensim, and Keras. To aid your understanding of popular ML algorithms, the course covers interesting and easy-to-follow examples such as news topic modeling and classification, spam email detection, stock price forecasting, and more. By the end of the course, you’ll have put together a broad picture of the ML ecosystem and will be well-versed with the best practices of applying ML techniques to make the most out of new opportunities.

Working in a hands-on learning environment, led by our Python Machine Learning instructor, students will learn about and explore:

* Exploit the power of Python to explore the world of data mining and data analytics
* Discover machine learning algorithms to solve complex challenges faced by data scientists today
* Use Python libraries such as TensorFlow and Keras to create smart cognitive actions for your projects

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the important concepts in machine learning and data science
* Use Python to explore the world of data mining and analytics
* Scale up model training using varied data complexities with Apache Spark
* Delve deep into text and NLP using Python libraries such NLTK and gensim
* Select and build an ML model and evaluate and optimize its performance
* Implement ML algorithms from scratch in Python, TensorFlow, and scikit-learn

**Audience & Pre-Requisites**

This course is geared for attendees with Apache knowledge who wish to Grasp machine learning concepts, techniques, and algorithms with the help of real-world examples using Python libraries such as TensorFlow and scikit-learn.

**Pre-Requisites:** Students should have

* Basic to python Skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Getting Started with Machine Learning and Python**

* Getting Started with Machine Learning and Python
* Defining machine learning and why we need it
* A very high-level overview of machine learning technology
* Core of machine learning – generalizing with data
* Preprocessing, exploration, and feature engineering
* Combining models
* Installing software and setting up

1. **Exploring the 20 Newsgroups Dataset with Text Analysis Techniques**

* Exploring the 20 Newsgroups Dataset with Text Analysis Techniques
* How computers understand language - NLP
* Picking up NLP basics while touring popular NLP libraries
* Getting the newsgroups data
* Exploring the newsgroups data
* Thinking about features for text data
* Visualizing the newsgroups data with t-SNE

1. **Mining the 20 Newsgroups Dataset with Clustering and Topic Modeling Algorithms**

* Mining the 20 Newsgroups Dataset with Clustering and Topic Modeling Algorithms
* Learning without guidance – unsupervised learning
* Clustering newsgroups data using k-means
* Discovering underlying topics in newsgroups
* Topic modeling using NMF
* Topic modeling using LDA

1. **Detecting Spam Email with Naive Bayes**

* Detecting Spam Email with Naive Bayes
* Getting started with classification
* Exploring Naïve Bayes
* Classification performance evaluation
* Model tuning and cross-validation

1. **Classifying Newsgroup Topics with Support Vector Machines**

* Classifying Newsgroup Topics with Support Vector Machines
* Finding separating boundary with support vector machines
* Classifying newsgroup topics with SVMs
* More example – fetal state classification on cardiotocography
* A further example – breast cancer classification using SVM with TensorFlow

1. **Predicting Online Ad Click-Through with Tree-Based Algorithms**

* Predicting Online Ad Click-Through with Tree-Based Algorithms
* Brief overview of advertising click-through prediction
* Getting started with two types of data – numerical and categorical
* Exploring decision tree from root to leaves
* Implementing a decision tree from scratch
* Predicting ad click-through with decision tree
* Ensembling decision trees – random forest

1. **Predicting Online Ad Click-Through with Logistic Regression**

* Predicting Online Ad Click-Through with Logistic Regression
* Converting categorical features to numerical – one-hot encoding and ordinal encoding
* Classifying data with logistic regression
* Training a logistic regression model
* Training on large datasets with online learning
* Handling multiclass classification
* Implementing logistic regression using TensorFlow
* Feature selection using random forest

1. **Scaling Up Prediction to Terabyte Click Logs**

* Scaling Up Prediction to Terabyte Click Logs
* Learning the essentials of Apache Spark
* Programming in PySpark
* Learning on massive click logs with Spark
* Feature engineering on categorical variables with Spark

1. **Stock Price Prediction with Regression Algorithms**

* Stock Price Prediction with Regression Algorithms
* Brief overview of the stock market and stock prices
* What is regression?
* Mining stock price data
* Estimating with linear regression
* Estimating with decision tree regression
* Estimating with support vector regression
* Estimating with neural networks
* Evaluating regression performance
* Predicting stock price with four regression algorithms

1. **Machine Learning Best Practices**

* Machine Learning Best Practices
* Machine learning solution workflow
* Best practices in the data preparation stage
* Best practices in the training sets generation stage
* Best practices in the model training, evaluation, and selection stage
* Best practices in the deployment and monitoring stage

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Bayesian Analysis with Python

**Course Snapshot**

* **Course:** Bayesian Analysis with Python
* **Duration:** 2 days
* **Skill-level**: Foundation-level Python Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to know Bayesian modeling with PyMC3 and exploratory analysis of Bayesian models with ArviZ
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Bayesian Analysis with Python is an introduction to the main concepts of applied Bayesian inference and its practical implementation in Python using PyMC3, a state-of-the-art probabilistic programming library, and ArviZ, a new library for exploratory analysis of Bayesian models. The main concepts of Bayesian statistics are covered using a practical and computational approach. Synthetic and real data sets are used to introduce several types of models, such as generalized linear models for regression and classification, mixture models, hierarchical models, and Gaussian processes, among others. By the end of the course, you will have a working knowledge of probabilistic modeling and you will be able to design and implement Bayesian models for your own data science problems. After reading the course you will be better prepared to delve into more advanced material or specialized statistical modeling if you need to.

Working in a hands-on learning environment, led by our Bayesian Analysis with Python instructor, students will learn about and explore:

* A step-by-step guide to conduct Bayesian data analyses using PyMC3 and ArviZ
* A modern, practical and computational approach to Bayesian statistical modeling
* A tutorial for Bayesian analysis and best practices with the help of sample problems and practice exercises.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Build probabilistic models using the Python library PyMC3
* Analyze probabilistic models with the help of ArviZ
* Acquire the skills required to sanity check models and modify them if necessary
* Understand the advantages and caveats of hierarchical models
* Find out how different models can be used to answer different data analysis questions
* Compare models and choose between alternative ones
* Discover how different models are unified from a probabilistic perspective
* Think probabilistically and benefit from the flexibility of the Bayesian framework

**Audience & Pre-Requisites**

This course is geared for attendees with Apache knowledge who wish to know Bayesian modeling with PyMC3 and exploratory analysis of Bayesian models with ArviZ

**Pre-Requisites:** Students should have

* Basic to python Skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Thinking Probabilistically**

* Thinking Probabilistically
* Statistics, models, and this course's approach
* Probability theory
* Single-parameter inference
* Communicating a Bayesian analysis
* Posterior predictive checks

1. **Programming Probabilistically**

* Programming Probabilistically
* Probabilistic programming
* PyMC3 primer
* Summarizing the posterior
* Gaussians all the way down
* Groups comparison
* Hierarchical models

1. **Modeling with Linear Regression**

* Modeling with Linear Regression
* Simple linear regression
* Robust linear regression
* Hierarchical linear regression
* Polynomial regression
* Multiple linear regression
* Variable variance

1. **Generalizing Linear Models**

* Generalizing Linear Models
* Generalized linear models
* Logistic regression
* Multiple logistic regression
* Poisson regression
* Robust logistic regression
* The GLM module

1. **Model Comparison**

* Model Comparison
* Posterior predictive checks
* Occam's razor – simplicity and accuracy
* Information criteria
* Bayes factors
* Regularizing priors
* WAIC in depth

1. **Mixture Models**

* Mixture Models
* Mixture models
* Finite mixture models
* Non-finite mixture model
* Continuous mixtures

1. **Gaussian Processes**

* Gaussian Processes
* Linear models and non-linear data
* Modeling functions
* Gaussian process regression
* Regression with spatial autocorrelation
* Gaussian process classification
* Cox processes

1. **Inference Engines**

* Inference Engines
* Inference engines
* Non-Markovian methods
* Markovian methods
* Diagnosing the samples

1. **Where To Go Next?**

* Where To Go Next?

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning for Data Mining

**Course Snapshot**

* **Course:** Machine Learning for Data Mining
* **Duration:** 1 days
* **Skill-level**: Foundation-level Machine Learning for Data Mining skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Get efficient in performing data mining and machine learning using IBM SPSS Modeler
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning (ML) combined with data mining can give you amazing results in your data mining work by empowering you with several ways to look at data. This course will help you improve your data mining techniques by using smart modeling techniques. This course will teach you how to implement ML algorithms and techniques in your data mining work. It will enable you to pair the best algorithms with the right tools and processes. You will learn how to identify patterns and make predictions with minimal human intervention. You will build different types of ML models, such as the neural network, the Support Vector Machines (SVMs), and the Decision tree. You will see how all these models works and what kind of data in the dataset they are suited for. You will learn how to combine the results of different models to improve accuracy. Topics such as removing noise and handling errors will give you an added edge in model building and optimization. By the end of this course, you will be able to build predictive models and extract information of interest from the dataset.

Working in a hands-on learning environment, led by our Machine Learning for Data Mining expert instructor, students will learn about and explore:

* Learn how to apply machine learning techniques in the field of data science
* Understand when to use different data mining techniques, how to set up different analyses, and how to interpret the results
* A step-by-step approach to improving model development and performance

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Hone your model-building skills and create the most accurate models
* Understand how predictive machine learning models work
* Prepare your data to acquire the best possible results
* Combine models to suit the requirements of different types of data
* Analyze single and multiple models and understand their combined results
* Derive worthwhile insights from your data using histograms and graphs

**Audience & Pre-Requisites**

This course is geared for attendees with Apache knowledge who wish to Get efficient in performing data mining and machine learning using IBM SPSS Modeler

**Pre-Requisites:** Students should have

* Basic to Python knowledge.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introducing Machine Learning Predictive Models**

* Introducing Machine Learning Predictive Models
* Characteristics of machine learning predictive models
* Types of machine learning predictive models
* Working with neural networks
* A sample neural network model

1. **Getting Started with Machine Learning**

* Getting Started with Machine Learning
* Demonstrating a neural network
* Support Vector Machines
* Demonstrating SVMs

1. **Understanding Models**

* Understanding Models
* Models
* Using graphs to interpret machine learning models
* Using statistics to interpret machine learning models
* Using decision trees to interpret machine learning models

1. **Improving Individual Models**

* Improving Individual Models
* Modifying model options
* Using a different model to improve results
* Removing noise to improve models
* Doing additional data preparation
* Balancing data

1. **Advanced Ways of Improving Models**

* Advanced Ways of Improving Models
* Combining models
* Using propensity scores
* Meta-level modeling
* Error modeling
* Boosting and bagging
* Predicting continuous outcomes

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning With Go

**Course Snapshot**

* **Course:** Machine Learning With Go
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning With Go skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Infuse an extra layer of intelligence into your Go applications with machine learning and AI
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

This popular Machine Learning With Go shows you how to overcome the common challenges of integrating analysis and machine learning code within an existing engineering organization. Machine Learning With Go, , will begin by helping you gain an understanding of how to gather, organize, and parse real-world data from a variety of sources. The course also provides absolute coverage in developing groundbreaking machine learning pipelines including predictive models, data visualizations, and statistical techniques. Up next, you will learn the thorough utilization of Golang libraries including golearn, gorgonia, gosl, hector, and mat64. You will discover the various TensorFlow capabilities, along with building simple neural networks and integrating them into machine learning models. You will also gain hands-on experience implementing essential machine learning techniques such as regression, classification, and clustering with the relevant Go packages. Furthermore, you will deep dive into the various Go tools that help you build deep neural networks. Lastly, you will become well versed with best practices for machine learning model tuning and optimization. By the end of the course, you will have a solid machine learning mindset and a powerful Go toolkit of techniques, packages, and example implementations.

Working in a hands-on learning environment, led by our Machine Learning With Go expert instructor, students will learn about and explore:

* Build simple, maintainable, and easy to deploy machine learning applications with popular Go packages
* Learn the statistics, algorithms, and techniques to implement machine learning
* Overcome the common challenges faced while deploying and scaling the machine learning workflows

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Become well versed with data processing, parsing, and cleaning using Go packages
* Learn to gather data from various sources and in various real-world formats
* Perform regression, classification, and image processing with neural networks
* Evaluate and detect anomalies in a time series model
* Understand common deep learning architectures to learn how each model is built
* Learn how to optimize, build, and scale machine learning workﬂows
* Discover the best practices for machine learning model tuning for successful deployments

**Audience & Pre-Requisites**

This course is geared for attendees with Apache knowledge who wish to Infuse an extra layer of intelligence into your Go applications with machine learning and AI

**Pre-Requisites:** Students should have

* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Gathering and Organizing Data**

* Gathering and Organizing Data
* Handling data – Gopher style
* Best practices for gathering and organizing data with Go
* CSV files
* Web scraping
* JSON
* SQL-like databases
* Caching
* Data versioning

1. **Matrices, Probability, and Statistics**

* Matrices, Probability, and Statistics
* Matrices and vectors
* Statistics
* Probability

1. **Evaluating and Validating**

* Evaluating and Validating
* Evaluating
* Validating

1. **Regression**

* Regression
* Understanding regression model jargon
* Linear regression
* Multiple linear regression
* Nonlinear and other types of regression

1. **Classification**

* Classification
* Understanding classification model jargon
* Logistic regression
* k-nearest neighbors
* Decision trees and random forests
* Naive Bayes

1. **Clustering**

* Clustering
* Understanding clustering model jargon
* Measuring distance or similarity
* Evaluating clustering techniques
* k-means clustering
* Other clustering techniques

1. **Time Series and Anomaly Detection**

* Time Series and Anomaly Detection
* Representing time series data in Go
* Understanding time series jargon
* Statistics related to time series
* Auto-regressive models for forecasting
* Auto-regressive moving averages and other time series models
* Anomaly detection

1. **Neural Networks**

* Neural Networks
* Understanding neural net jargon
* Building a simple neural network
* Utilizing the simple neural network

1. **Deep Learning**

* Deep Learning
* Deep learning techniques and jargon
* Deep learning with Go

1. **Deploying and Distributing Analyses and Models**

* Deploying and Distributing Analyses and Models
* Running models reliably on remote machines
* Building a scalable and reproducible machine learning pipeline

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with R

**Course Snapshot**

* **Course:** Machine Learning with R
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Solve real-world data problems with R and machine learning
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning, at its core, is concerned with transforming data into actionable knowledge. R offers a powerful set of machine learning methods to quickly and easily gain insight from your data. Machine Learning with R, This Edition provides a hands-on, readable guide to applying machine learning to real-world problems. Whether you are an experienced R user or new to the language, course teaches you everything you need to uncover key insights, make new predictions, and visualize your findings. This updates the classic R data science course with newer and better libraries, advice on ethical and bias issues in machine learning, and an introduction to deep learning. Find powerful new insights in your data; discover machine learning with R..

Working in a hands-on learning environment, led by our Machine Learning with R expert instructor, students will learn about and explore:

* bestselling, widely acclaimed R machine learning course, updated and improved for R 3.5 and beyond
* Harness the power of R to build flexible, effective, and transparent machine learning models
* Learn quickly with a clear, hands-on guide by experienced machine learning teacher and practitioner

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Discover the origins of machine learning and how exactly a computer learns by example
* Prepare your data for machine learning work with the R programming language
* Classify important outcomes using nearest neighbor and Bayesian methods
* Predict future events using decision trees, rules, and support vector machines
* Forecast numeric data and estimate financial values using regression methods
* Model complex processes with artificial neural networks — the basis of deep learning
* Avoid bias in machine learning models
* Evaluate your models and improve their performance
* Connect R to SQL databases and emerging big data technologies such as Spark, H2O, and TensorFlow

**Audience & Pre-Requisites**

This course is geared for attendees who wish to Solve real-world data problems with R and machine learning

**Pre-Requisites:** Students should have

* Basic to ML Skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introducing Machine Learning**

* Introducing Machine Learning
* The origins of machine learning
* Uses and abuses of machine learning
* How machines learn
* Machine learning in practice
* Machine learning with R

1. **Managing and Understanding Data**

* Managing and Understanding Data
* R data structures
* Managing data with R
* Exploring and understanding data

1. **Lazy Learning – Classification Using Nearest Neighbors**

* Lazy Learning – Classification Using Nearest Neighbors
* Understanding nearest neighbor classification
* Example – diagnosing breast cancer with the k-NN algorithm

1. **Probabilistic Learning – Classification Using Naive Bayes**

* Probabilistic Learning – Classification Using Naive Bayes
* Understanding Naive Bayes
* Example – filtering mobile phone spam with the Naive Bayes algorithm

1. **Divide and Conquer – Classification Using Decision Trees and Rules**

* Divide and Conquer – Classification Using Decision Trees and Rules
* Understanding decision trees
* Example – identifying risky bank loans using C5.0 decision trees
* Understanding classification rules
* Example – identifying poisonous mushrooms with rule learners

1. **Forecasting Numeric Data – Regression Methods**

* Forecasting Numeric Data – Regression Methods
* Understanding regression
* Example – predicting medical expenses using linear regression
* Understanding regression trees and model trees
* Example – estimating the quality of wines with regression trees and model trees

1. **Black Box Methods – Neural Networks and Support Vector Machines**

* Black Box Methods – Neural Networks and Support Vector Machines
* Understanding neural networks
* Example – modeling the strength of concrete with ANNs
* Understanding support vector machines
* Example – performing OCR with SVMs

1. **Finding Patterns – Market Basket Analysis Using Association Rules**

* Finding Patterns – Market Basket Analysis Using Association Rules
* Understanding association rules
* Example – identifying frequently purchased groceries with association rules

1. **Finding Groups of Data – Clustering with k-means**

* Finding Groups of Data – Clustering with k-means
* Understanding clustering
* Finding teen market segments using k-means clustering

1. **Evaluating Model Performance**

* Evaluating Model Performance
* Measuring performance for classification
* Estimating future performance

1. **Improving Model Performance**

* Improving Model Performance
* Tuning stock models for better performance
* Improving model performance with meta-learning

1. **Specialized Machine Learning Topics**

* Specialized Machine Learning Topics
* Managing and preparing real-world data
* Working with online data and services
* Working with domain-specific data
* Improving the performance of R

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with Scala

**Course Snapshot**

* **Course:** Machine Learning with Scala
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning with Scala skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to know Supervised and unsupervised machine learning made easy in Scala with this quick-start guide.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Scala is a highly scalable integration of object-oriented nature and functional programming concepts that make it easy to build scalable and complex big data applications. This course is a handy guide for machine learning developers and data scientists who want to develop and train effective machine learning models in Scala. The course starts with an introduction to machine learning, while covering deep learning and machine learning basics. It then explains how to use Scala-based ML libraries to solve classification and regression problems using linear regression, generalized linear regression, logistic regression, support vector machine, and Naïve Bayes algorithms. It also covers tree-based ensemble techniques for solving both classification and regression problems. Moving ahead, it covers unsupervised learning techniques, such as dimensionality reduction, clustering, and recommender systems. Finally, it provides a brief overview of deep learning using a real-life example in Scala.

Working in a hands-on learning environment, led by our Machine Learning with Scala expert instructor, students will learn about and explore:

* Construct and deploy machine learning systems that learn from your data and give accurate predictions
* Unleash the power of Spark ML along with popular machine learning algorithms to solve complex tasks in Scala.
* Solve hands-on problems by combining popular neural network architectures such as LSTM and CNN using Scala with DeepLearning4j library

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get acquainted with JVM-based machine learning libraries for Scala such as Spark ML and Deeplearning4j
* Learn RDDs, DataFrame, and Spark SQL for analyzing structured and unstructured data
* Understand supervised and unsupervised learning techniques with best practices and pitfalls
* Learn classification and regression analysis with linear regression, logistic regression, Naïve Bayes, support vector machine, and tree-based ensemble techniques
* Learn effective ways of clustering analysis with dimensionality reduction techniques
* Learn recommender systems with collaborative filtering approach
* Delve into deep learning and neural network architectures

**Audience & Pre-Requisites**

This course is geared for attendees with Apache knowledge who wish to know Supervised and unsupervised machine learning made easy in Scala with this quick-start guide.

**Pre-Requisites:** Students should have

* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction to Machine Learning with Scala**

* Introduction to Machine Learning with Scala
* Technical requirements
* Overview of ML
* ML tasks
* Overview of Scala
* ML libraries in Scala
* Getting started learning

1. **Scala for Regression Analysis**

* Scala for Regression Analysis
* Technical requirements
* An overview of regression analysis
* Regression analysis algorithms
* Learning regression analysis through examples
* Linear regression
* Generalized linear regression (GLR)
* Hyperparameter tuning and cross-validation

1. **Scala for Learning Classification**

* Scala for Learning Classification
* Technical requirements
* Overview of classification
* Developing predictive models for churn
* LR for churn prediction
* NB for churn prediction
* SVM for churn prediction

1. **Scala for Tree-Based Ensemble Techniques**

* Scala for Tree-Based Ensemble Techniques
* Technical requirements
* Decision trees and tree ensembles
* Decision trees for supervised learning
* Gradient boosted trees for supervised learning
* Random forest for supervised learning

1. **Scala for Dimensionality Reduction and Clustering**

* Scala for Dimensionality Reduction and Clustering
* Technical requirements
* Overview of unsupervised learning
* Clustering analysis through examples
* Dimensionality reduction

1. **Scala for Recommender System**

* Scala for Recommender System
* Technical requirements
* Overview of recommendation systems
* Model-based course recommendation system

1. **Introduction to Deep Learning with Scala**

* Introduction to Deep Learning with Scala
* Technical requirements
* DL versus ML
* DL and ANNs
* Neural network architectures
* DL frameworks
* Getting started with learning

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Q-Learning with Python

**Course Snapshot**

* **Course:** Q-Learning with Python
* **Duration:** 2 days
* **Skill-level**: Foundation-level Q-Learning with Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to get the Leverage power of reward-based training for your deep learning models with Python
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Q-learning is a machine learning algorithm used to solve optimization problems in artificial intelligence (AI). It is one of the most popular fields of study among AI researchers. This course starts off by introducing you to reinforcement learning and Q-learning, in addition to helping you become familiar with OpenAI Gym as well as libraries such as Keras and TensorFlow. A few lessons into the course, you will gain insights into model-free Q-learning and use deep Q-networks and double deep Q-networks to solve complex problems. This course will guide you in exploring use cases such as self-driving vehicles and OpenAI Gym’s CartPole problem. You will also learn how to tune and optimize Q-networks and their hyperparameters. As you progress, you will understand the reinforcement learning approach to solving real-world problems. You will also explore how to use Q-learning and related algorithms in scientific research. Toward the end, you’ll gain insight into what’s in store for reinforcement learning. By the end of this course, you will be equipped with the skills you need to solve reinforcement learning problems using Q-learning algorithms with OpenAI Gym, Keras, and TensorFlow.

Working in a hands-on learning environment, led by our Q-Learning with Python expert instructor, students will learn about and explore:

* Understand Q-learning algorithms to train neural networks using Markov Decision Process (MDP)
* Study practical deep reinforcement learning using Q-Networks
* Explore state-based unsupervised learning for machine learning models

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Explore the fundamentals of reinforcement learning and the state-action-reward process
* Understand Markov Decision Processes
* Get well-versed with libraries such as Keras, and TensorFlow
* Create and deploy model-free learning and deep Q-learning agents with TensorFlow, Keras, and OpenAI Gym
* Choose and optimize a Q-network’s learning parameters and fine-tune its performance
* Discover real-world applications and use cases of Q-learning

**Audience & Pre-Requisites**

This course is geared for attendees with Apache knowledge who wish to know the Leverage power of reward-based training for your deep learning models with Python.

**Pre-Requisites:** Students should have

* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Brushing Up on Reinforcement Learning Concepts**

* Brushing Up on Reinforcement Learning Concepts
* What is RL?
* States, actions, and rewards
* Key concepts in RL
* SARSA versus Q-learning – on-policy or off?

1. **Getting Started with the Q-Learning Algorithm**

* Getting Started with the Q-Learning Algorithm
* Technical requirements
* Demystifying MDPs
* Your Q-learning agent in its environment
* Fine-tuning your model – learning, discount, and exploration rates
* MABP – a classic exploration versus exploitation problem
* Optimal versus safe paths – revisiting SARSA

1. **Setting Up Your First Environment with OpenAI Gym**

* Setting Up Your First Environment with OpenAI Gym
* Technical requirements
* Getting started with OpenAI Gym
* Exploring the Taxi-v2 environment
* Creating a baseline agent

1. **Teaching a Smartcab to Drive Using Q-Learning**

* Teaching a Smartcab to Drive Using Q-Learning
* Technical requirements
* Getting to know your learning agent
* Implementing your agent
* The learning parameters – alpha, gamma, and epsilon
* Model-tuning and tracking your agent's long-term performance

1. **Building Q-Networks with TensorFlow**

* Building Q-Networks with TensorFlow
* Technical requirements
* A brief overview of neural networks
* Taking a closer look
* Implementing a neural network with NumPy
* Neural networks and Q-learning
* Building your first Q-network

1. **Digging Deeper into Deep Q-Networks with Keras and TensorFlow**

* Digging Deeper into Deep Q-Networks with Keras and TensorFlow
* Technical requirements
* Introducing CartPole-v1
* Getting started with the CartPole task
* Building a DQN to solve the CartPole problem
* Testing and results
* Adding in experience replay
* Building further on DQNs

1. **Decoupling Exploration and** **Exploitation in Multi-Armed Bandits**

* Decoupling Exploration and Exploitation in Multi-Armed Bandits
* Technical requirements
* Probability distributions and ongoing knowledge
* Revisiting a simple bandit problem
* Multi-armed bandit strategy overview
* Contextual bandits and state diagrams
* Thompson sampling and the Bayesian control rule
* Solving a multi-armed bandit problem in Python – user advertisement clicks
* Multi-armed bandits in experimental design

1. **Further Q-Learning Research and Future Projects**

* Further Q-Learning Research and Future Projects
* Google's DeepMind and the future of Q-learning
* OpenAI Gym and RL research
* More OpenAI Gym environments
* Contextual bandits and probability distributions

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# C# Machine Learning Projects

**Course Snapshot**

* **Course:** C# Machine Learning Projects
* **Duration:** 3 days
* **Skill-level**: Foundation-level C# Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Power C# and .NET applications with exciting machine learning models and modular projects
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning is applied in almost all kinds of real-world surroundings and industries, right from medicine to advertising; from finance to scientifc research. This course will help you learn how to choose a model for your problem, how to evaluate the performance of your models, and how you can use C# to build machine learning models for your future projects. You will get an overview of the machine learning systems and how you, as a C# and .NET developer, can apply your existing knowledge to the wide gamut of intelligent applications, all through a project-based approach. You will start by setting up your C# environment for machine learning with the required packages, Accord.NET, LiveCharts, and Deedle. We will then take you right from building classifcation models for spam email fltering and applying NLP techniques to Twitter sentiment analysis, to time-series and regression analysis for forecasting foreign exchange rates and house prices, as well as drawing insights on customer segments in e-commerce. You will then build a recommendation model for music genre recommendation and an image recognition model for handwritten digits. Lastly, you will learn how to detect anomalies in network and credit card transaction data for cyber attack and credit card fraud detections. By the end of this course, you will be putting your skills in practice and implementing your machine learning knowledge in real projects.

Working in a hands-on learning environment, led by our C# Machine Learning expert instructor, students will learn about and explore:

* Produce classification, regression, association, and clustering models
* Expand your understanding of machine learning and C#
* Get to grips with C# packages such as Accord.net, LiveCharts, and Deedle

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Set up the C# environment for machine learning with required packages
* Build classification models for spam email filtering
* Get to grips with feature engineering using NLP techniques for Twitter sentiment analysis
* Forecast foreign exchange rates using continuous and time-series data
* Make a recommendation model for music genre recommendation
* Familiarize yourself with munging image data and Neural Network models for handwritten-digit recognition
* Use Principal Component Analysis (PCA) for cyber attack detection
* One-Class Support Vector Machine for credit card fraud detection

**Audience & Pre-Requisites**

This course is geared for attendees who wish to Power C# and .NET applications with exciting machine learning models and modular projects

**Pre-Requisites:** Students should have

* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Basics of Machine Learning Modeling**

* Basics of Machine Learning Modeling
* Key ML tasks and applications
* Steps in building ML models
* Setting up a C# environment for ML

1. **Spam Email Filtering**

* Spam Email Filtering
* Problem definition for the spam email filtering project
* Data preparation
* Email data analysis
* Feature engineering for email data
* Logistic regression versus Naive Bayes for email spam filtering
* Classification model validations

1. **Twitter Sentiment Analysis**

* Twitter Sentiment Analysis
* Setting up the environment
* Problem definition for Twitter sentiment analysis
* Data preparation using Stanford CoreNLP
* Data analysis using lemmas as tokens
* Feature engineering using lemmatization and emoticons
* Naive Bayes versus random forest
* Model validations – ROC curve and AUC

1. **Foreign Exchange Rate Forecast**

* Foreign Exchange Rate Forecast
* Problem definition
* Data preparation
* Time series data analysis
* Feature engineering
* Linear regression versus SVM
* Model validations

1. **Fair Value of House and Property**

* Fair Value of House and Property
* Problem definition
* Categorical versus continuous variables
* Feature engineering and encoding
* Linear regression versus SVM with kernels
* Model validations

1. **Customer Segmentation**

* Customer Segmentation
* Problem definition
* Data analysis for the online retail dataset
* Feature engineering and data aggregation
* Unsupervised learning – k-means clustering
* Clustering model validations using the Silhouette Coefficient

1. **Music Genre Recommendation**

* Music Genre Recommendation
* Problem definition
* Data analysis for the audio features dataset
* ML models for music genre classification
* Ensembling base learning models
* Evaluating recommendation/rank-ordering models

1. **Handwritten Digit Recognition**

* Handwritten Digit Recognition
* Problem definition
* Data analysis for the image dataset
* Feature engineering and dimensionality reduction
* ML models for handwritten digit recognition
* Evaluating multi-class classification models

1. **Cyber Attack Detection**

* Cyber Attack Detection
* Problem definition
* Data analysis for internet traffic data
* Feature engineering and PCA
* Principal component classifier for anomaly detection
* Evaluating anomaly detection models

1. **Credit Card Fraud Detection**

* Credit Card Fraud Detection
* Problem definition
* Data analysis for anonymized credit card data
* Feature engineering and PCA
* One-class SVM versus PCC
* Evaluating anomaly detection models

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with C#

**Course Snapshot**

* **Course:** Machine Learning with C#
* **Duration:** 3 days
* **Skill-level**: Foundation-level C# Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Explore supervised and unsupervised learning techniques and add smart features to your applications
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The necessity for machine learning is everywhere, and most production enterprise applications are written in C# using tools such as Visual Studio, SQL Server, and Microsoft Azur2e. Hands-On Machine Learning with C# uniquely blends together an understanding of various machine learning concepts, techniques of machine learning, and various available machine learning tools through which users can add intelligent features. These tools include image and motion detection, Bayes intuition, and deep learning, to C# .NET applications. Using this course, you will learn to implement supervised and unsupervised learning algorithms and will be better equipped to create excellent predictive models. In addition, you will learn both supervised and unsupervised forms of regression, mainly logistic and linear regression, in depth. Next, you will use the nuML machine learning framework to learn how to create a simple decision tree. In the concluding lessons, you will use the Accord.Net machine learning framework to learn sequence recognition of handwritten numbers using dynamic time warping. We will also cover advanced concepts such as artificial neural networks, autoencoders, and reinforcement learning. By the end of this course, you will have developed a machine learning mindset and will be able to leverage C# tools, techniques, and packages to build smart, predictive, and real-world business applications.

Working in a hands-on learning environment, led by our C# Machine Learning expert instructor, students will learn about and explore:

* Leverage machine learning techniques to build real-world applications
* Use the Accord.NET machine learning framework for reinforcement learning
* Implement machine learning techniques using Accord, nuML, and Encog

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn to parameterize a probabilistic problem
* Use Naive Bayes to visually plot and analyze data
* Plot a text-based representation of a decision tree using nuML
* Use the Accord.NET machine learning framework for associative rule-based learning
* Develop machine learning algorithms utilizing fuzzy logic
* Explore support vector machines for image recognition
* Understand dynamic time warping for sequence recognition

**Audience & Pre-Requisites**

This course is geared for attendees who wish to Explore supervised and unsupervised learning techniques and add smart features to your applications

**Pre-Requisites:** Students should have

* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Machine Learning Basics**

* Machine Learning Basics
* Introduction to machine learning
* Data mining
* Artificial Intelligence
* Bio-AI
* Deep learning
* Probability and statistics
* Approaching your machine learning project
* Iris dataset
* Supervised learning
* Unsupervised learning
* Reinforcement learning
* Build, buy, or open source

1. **ReflectInsight – Real-Time Monitoring**

* ReflectInsight – Real-Time Monitoring
* Router
* Log Viewer
* Live Viewer

1. **Bayes Intuition – Solving the Hit and Run Mystery and Performing Data Analysis**

* Bayes Intuition – Solving the Hit and Run Mystery and Performing Data Analysis
* Overviewing Bayes' theorem
* Overviewing Naive Bayes and plotting data

1. **Risk versus Reward – Reinforcement Learning**

* Risk versus Reward – Reinforcement Learning
* Overviewing reinforcement learning
* Types of learning
* Q-learning
* SARSA
* Running our application
* Tower of Hanoi

1. **Fuzzy Logic – Navigating the** **Obstacle Course**

* Fuzzy Logic – Navigating the Obstacle Course
* Fuzzy logic

1. **Color Blending – Self-Organizing Maps and Elastic Neural Networks**

* Color Blending – Self-Organizing Maps and Elastic Neural Networks
* Under the hood of an SOM

1. **Facial and Motion Detection – Imaging Filters**

* Facial and Motion Detection – Imaging Filters
* Facial detection
* Motion detection

1. **Encyclopedias and Neurons – Traveling Salesman Problem**

* Encyclopedias and Neurons – Traveling Salesman Problem
* Traveling salesman problem
* Learning rate parameter

1. **Should I Take the Job – Decision Trees in Action**

* Should I Take the Job – Decision Trees in Action
* Decision tree
* Should I take the job?
* numl
* Accord.NET decision trees

1. **Deep Belief – Deep Networks and Dreaming**

* Deep Belief – Deep Networks and Dreaming
* Restricted Boltzmann Machines
* What does a computer dream?

1. **Microbenchmarking and Activation Functions**

* Microbenchmarking and Activation Functions
* Visual activation function plotting

1. **Intuitive Deep Learning in C# .NET**

* Intuitive Deep Learning in C# .NET
* What is deep learning?
* The Kelp.Net Framework

1. **Quantum Computing – The Future**

* Quantum Computing – The Future
* Superposition
* Teleportation

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with Microsoft Excel 2019

**Course Snapshot**

* **Course:** Machine Learning with Microsoft Excel 2019
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning with Microsoft Excel 2019 skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants a practical guide to getting the most out of Excel, using it for data preparation, applying machine learning models (including cloud services) and understanding the outcome of the data analysis.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

We have made huge progress in teaching computers to perform difficult tasks, especially those that are repetitive and time-consuming for humans. Excel users, of all levels, can feel left behind by this innovation wave. The truth is that a large amount of the work needed to develop and use a machine learning model can be done in Excel. The course starts by giving a general introduction to machine learning, making every concept clear and understandable. Then, it shows every step of a machine learning project, from data collection, reading from different data sources, developing models, and visualizing the results using Excel features and offerings. In every lesson, there are several examples and hands-on exercises that will show the reader how to combine Excel functions, add-ins, and connections to databases and to cloud services to reach the desired goal: building a full data analysis flow. Different machine learning models are shown, tailored to the type of data to be analyzed.

Working in a hands-on learning environment, led by our Machine Learning with Microsoft Excel 2019 expert instructor, students will learn about and explore:

* Use Microsoft's product Excel to build advanced forecasting models using varied examples
* Cover range of machine learning tasks such as data mining, data analytics, smart visualization, and more
* Derive data-driven techniques using Excel plugins and APIs without much code required

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use Excel to preview and cleanse datasets
* Understand correlations between variables and optimize the input to machine learning models
* Use and evaluate different machine learning models from Excel
* Understand the use of different visualizations
* Learn the basic concepts and calculations to understand how artificial neural networks work
* Learn how to connect Excel to the Microsoft Azure cloud
* Get beyond proof of concepts and build fully functional data analysis flows

**Audience & Pre-Requisites**

This course is geared for attendees with Machine Learning with Microsoft Excel 2019 who wish some advance use cases using Automated Machine Learning, and artificial neural network, which simplifies the analysis task and represents the future of machine learning.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Machine Learning with Microsoft Excel 2019 knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Implementing Machine Learning Algorithms**

* Technical requirements
* Understanding learning and models
* Focusing on model features
* Studying machine learning models in practice
* Comparing underfitting and overfitting
* Evaluating models

1. **Hands-On Examples of Machine Learning Models**

* Technical requirements
* Understanding supervised learning with multiple linear regression
* Understanding supervised learning with decision trees
* Understanding unsupervised learning with clustering

1. **Importing Data into Excel from Different Data Sources**

* Technical requirements
* Importing data from a text file
* Importing data from another Excel workbook
* Importing data from a web page
* Importing data from Facebook
* Importing data from a JSON file
* Importing data from a database

1. **Data Cleansing and Preliminary Data Analysis**

* Technical requirements
* Cleansing data
* Visualizing data for preliminary analysis
* Understanding unbalanced datasets

1. **Correlations and the Importance of Variables**

* Technical requirements
* Building a scatter diagram
* Calculating the covariance
* Calculating the Pearson's coefficient of correlation
* Studying the Spearman's correlation
* Understanding least squares
* Focusing on feature selection

1. **Data Mining Models in Excel Hands-On Examples**

* Technical requirements
* Learning by example – Market Basket Analysis
* Learning by example – Customer Cohort Analysis

1. **Implementing Time Series**

* Technical requirements
* Modeling and visualizing time series
* Forecasting time series automatically in Excel
* Studying the stationarity of a time series

1. **Visualizing Data in Diagrams, Histograms, and Maps**

* Technical requirements
* Showing basic comparisons and relationships between variables
* Building data distributions using histograms
* Representing geographical distribution of data in maps
* Showing data that changes over time

1. **Artificial Neural Networks**

* Technical requirements
* Introducing the perceptron – the simplest type of neural network
* Building a deep network
* Understanding the backpropagation algorithm

1. **Azure and Excel - Machine Learning in the Cloud**

* Technical requirements
* Introducing the Azure Cloud
* Using AMLS for free – a step-by-step guide
* Loading your data into AMLS
* Creating and running an experiment in AMLS

1. **The Future of Machine Learning**

* Automatic data analysis flows
* Automated machine learning

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with JavaScript

**Course Snapshot**

* **Course:** Machine Learning with JavaScript
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning with JavaScript skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants a definitive guide to creating an intelligent web application with the best of machine learning and JavaScript.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In over 20 years of existence, JavaScript has been pushing beyond the boundaries of web evolution with proven existence on servers, embedded devices, Smart TVs, IoT, Smart Cars, and more. Today, with the added advantage of machine learning research and support for JS libraries, JavaScript makes your browsers smarter than ever with the ability to learn patterns and reproduce them to become a part of innovative products and applications. Machine Learning with JavaScript presents various avenues of machine learning in a practical and objective way, and helps implement them using the JavaScript language. Predicting behaviors, analyzing feelings, grouping data, and building neural models are some of the skills you will build from this course. You will learn how to train your machine learning models and work with different kinds of data. During this journey, you will come across use cases such as face detection, spam filtering, recommendation systems, character recognition, and more. Moreover, you will learn how to work with deep neural networks and guide your applications to gain insights from data. By the end of this course, you'll have gained hands-on knowledge on evaluating and implementing the right model, along with choosing from different JS libraries, such as NaturalNode, brain, harthur, classifier, and many more to design smarter applications.

Working in a hands-on learning environment, led by our Machine Learning with JavaScript expert instructor, students will learn about and explore:

* Solve complex computational problems in browser with JavaScript
* Teach your browser how to learn from rules using the power of machine learning
* Understand discoveries on web interface and API in machine learning

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get an overview of state-of-the-art machine learning
* Understand the pre-processing of data handling, cleaning, and preparation
* Learn Mining and Pattern Extraction with JavaScript
* Build your own model for classification, clustering, and prediction
* Identify the most appropriate model for each type of problem
* Apply machine learning techniques to real-world applications
* Learn how JavaScript can be a powerful language for machine learning

**Audience & Pre-Requisites**

This course is geared for attendees who wish a definitive guide to creating an intelligent web application with the best of machine learning and JavaScript.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Exploring the Potential of JavaScript**

* Exploring the Potential of JavaScript
* Why JavaScript?
* Why machine learning, why now?
* Advantages and challenges of JavaScript
* The CommonJS initiative
* Node.js
* TypeScript language
* Improvements in ES6
* Preparing the development environment

1. **Data Exploration**

* Data Exploration
* An overview
* Feature identification
* Cleaning and preparing data

1. **Tour of Machine Learning Algorithms**

* Tour of Machine Learning Algorithms
* Introduction to machine learning
* Types of learning
* Categories of algorithms

1. **Grouping with Clustering Algorithms**

* Grouping with Clustering Algorithms
* Average and distance
* Writing the k-means algorithm
* Example 1 – k-means on simple 2D data
* Example 2 – 3D data
* k-means where k is unknown

1. **Classification Algorithms**

* Classification Algorithms
* k-Nearest Neighbor
* Naive Bayes classifier
* Support Vector Machine
* Random forest

1. **Association Rule Algorithms**

* Association Rule Algorithms
* The mathematical perspective
* The algorithmic perspective
* Association rule applications
* Example – retail data

1. **Forecasting with Regression Algorithms**

* Forecasting with Regression Algorithms
* Regression versus classification
* Regression basics
* Example 1 – linear regression
* Example 2 – exponential regression
* Example 3 – polynomial regression
* Other time-series analysis techniques

1. **Artificial Neural Network Algorithms**

* Artificial Neural Network Algorithms
* Conceptual overview of neural networks
* Backpropagation training
* Example - XOR in TensorFlow.js

1. **Deep Neural Networks**

* Deep Neural Networks
* Convolutional Neural Networks
* Recurrent neural networks

1. **Natural Language Processing in Practice**

* Natural Language Processing in Practice
* String distance
* Term frequency - inverse document frequency
* Tokenizing
* Stemming
* Phonetics
* Part of speech tagging
* Word embedding and neural networks

1. **Using Machine Learning in Real-Time Applications**

* Using Machine Learning in Real-Time Applications
* Serializing models
* Data pipelines

1. **Choosing the Best Algorithm for Your Application**

* Choosing the Best Algorithm for Your Application
* Mode of learning
* The task at hand
* Format, form, input, and output
* Available resources
* When it goes wrong
* Combining models

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Ensemble Learning with Python

**Course Snapshot**

* **Course:** Ensemble Learning with Python
* **Duration:** 3 days
* **Skill-level**: Foundation-level Ensemble Learning with Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Combine popular machine learning techniques to create ensemble models using Python.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Ensembling is a technique of combining two or more similar or dissimilar machine learning algorithms to create a model that delivers superior predictive power. This course will demonstrate how you can use a variety of weak algorithms to make a strong predictive model. With its hands-on approach, you'll not only get up to speed with the basic theory but also the application of different ensemble learning techniques. Using examples and real-world datasets, you'll be able to produce better machine learning models to solve supervised learning problems such as classification and regression. In addition to this, you'll go on to leverage ensemble learning techniques such as clustering to produce unsupervised machine learning models. As you progress, the lessons will cover different machine learning algorithms that are widely used in the practical world to make predictions and classifications. You'll even get to grips with the use of Python libraries such as scikit-learn and Keras for implementing different ensemble models. By the end of this course, you will be well-versed in ensemble learning, and have the skills you need to understand which ensemble method is required for which problem, and successfully implement them in real-world scenarios.

Working in a hands-on learning environment, led by our Python expert instructor, students will learn about and explore:

* Implement ensemble models using algorithms such as random forests and AdaBoost
* Apply boosting, bagging, and stacking ensemble methods to improve the prediction accuracy of your model
* Explore real-world data sets and practical examples coded in scikit-learn and Keras

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Implement ensemble methods to generate models with high accuracy
* Overcome challenges such as bias and variance
* Explore machine learning algorithms to evaluate model performance
* Understand how to construct, evaluate, and apply ensemble models
* Analyze tweets in real time using Twitter's streaming API
* Use Keras to build an ensemble of neural networks for the MovieLens dataset

**Audience & Pre-Requisites**

This course is geared for attendees wants to Combine popular machine learning techniques to create ensemble models using Python

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Python programming knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **A Machine Learning Refresher**

* A Machine Learning Refresher
* Technical requirements
* Learning from data
* Supervised and unsupervised learning
* Performance measures
* Machine learning algorithms

1. **Getting Started with Ensemble Learning**

* Getting Started with Ensemble Learning
* Technical requirements
* Bias, variance, and the trade-off
* Ensemble learning
* Difficulties in ensemble learning

1. **Voting**

* Voting
* Technical requirements
* Hard and soft voting
* Python implementation
* Using scikit-learn

1. **Stacking**

* Stacking
* Technical requirements
* Meta-learning
* Deciding on an ensemble's composition
* Python implementation

1. **Bagging**

* Bagging
* Technical requirements
* Bootstrapping
* Bagging
* Python implementation
* Using scikit-learn

1. **Boosting**

* Boosting
* Technical requirements
* AdaBoost
* Gradient boosting
* Using scikit-learn
* XGBoost

1. **Random Forests**

* Random Forests
* Technical requirements
* Understanding random forest trees
* Creating forests
* Using scikit-learn

1. **Clustering**

* Clustering
* Technical requirements
* Consensus clustering
* Using OpenEnsembles

1. **Classifying Fraudulent Transactions**

* Classifying Fraudulent Transactions
* Technical requirements
* Getting familiar with the dataset
* Exploratory analysis
* Voting
* Stacking
* Bagging
* Boosting
* Using random forests
* Comparative analysis of ensembles

1. **Predicting Bitcoin Prices**

* Predicting Bitcoin Prices
* Technical requirements
* Time series data
* Voting
* Stacking
* Bagging
* Boosting
* Random forests

1. **Evaluating Sentiment on Twitter**

* Evaluating Sentiment on Twitter
* Technical requirements
* Sentiment analysis tools
* Getting Twitter data
* Creating a model
* Classifying tweets in real time

1. **Recommending Movies with Keras**

* Recommending Movies with Keras
* Technical requirements
* Demystifying recommendation systems
* Neural recommendation systems
* Using Keras for movie recommendations

1. **Clustering World Happiness**

* Clustering World Happiness
* Technical requirements
* Understanding the World Happiness Report
* Creating the ensemble
* Gaining insights

Student Materials: Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work

# Machine Learning Fundamentals

**Course Snapshot**

* **Course:** Machine Learning Fundamentals
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning Fundamentals skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to know the features of scikit-learn and Python, build machine learning algorithms that optimize the programming process and take application performance to a whole new level.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

As machine learning algorithms become popular, new tools that optimize these algorithms are also developed. Machine Learning Fundamentals explains you how to use the syntax of scikit-learn. You'll study the difference between supervised and unsupervised models, as well as the importance of choosing the appropriate algorithm for each dataset. You'll apply unsupervised clustering algorithms over real-world datasets, to discover patterns and profiles, and explore the process to solve an unsupervised machine learning problem. The focus of the course then shifts to supervised learning algorithms. You'll learn to implement different supervised algorithms and develop neural network structures using the scikit-learn package. You'll also learn how to perform coherent result analysis to improve the performance of the algorithm by tuning hyperparameters. By the end of this course, you will have gain all the skills required to start programming machine learning algorithms.

Working in a hands-on learning environment, led by our Machine Learning Fundamentals expert instructor, students will learn about and explore:

* Explore scikit-learn uniform API and its application into any type of model
* Understand the difference between supervised and unsupervised models
* Learn the usage of machine learning through real-world examples

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the importance of data representation
* Gain insights into the differences between supervised and unsupervised models
* Explore data using the Matplotlib library
* Study popular algorithms, such as k-means, Mean-Shift, and DBSCAN
* Measure model performance through different metrics
* Implement a confusion matrix using scikit-learn
* Study popular algorithms, such as Naïve-Bayes, Decision Tree, and SVM
* Perform error analysis to improve the performance of the model
* Learn to build a comprehensive machine learning program

**Audience & Pre-Requisites**

This course is geared for attendees wants to know the features of scikit-learn and Python, build machine learning algorithms that optimize the programming process and take application performance to a whole new level

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Python programming knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction to Scikit-Learn**

* Introduction to Scikit-Learn
* Introduction
* Scikit-Learn
* Data Representation
* Data Preprocessing
* Scikit-Learn API
* Supervised and Unsupervised Learning

1. **Unsupervised Learning: Real-Life Applications**

* Unsupervised Learning: Real-Life Applications
* Introduction
* Clustering
* Exploring a Dataset: Wholesale Customers Dataset
* Data Visualization
* k-means Algorithm
* Mean-Shift Algorithm
* DBSCAN Algorithm
* Evaluating the Performance of Clusters

1. **Supervised Learning: Key Steps**

* Supervised Learning: Key Steps
* Introduction
* Model Validation and Testing
* Evaluation Metrics
* Error Analysis

1. **Supervised Learning Algorithms: Predict Annual Income**

* Supervised Learning Algorithms: Predict Annual Income
* Introduction
* Exploring the Dataset
* Naïve Bayes Algorithm
* Decision Tree Algorithm
* Support Vector Machine Algorithm
* Error Analysis

1. **Artificial Neural Networks:** **Predict Annual Income**

* Artificial Neural Networks: Predict Annual Income
* Introduction
* Artificial Neural Networks
* Applying an Artificial Neural Network
* Performance Analysis

1. **Building Your Own Program**

* Building Your Own Program
* Introduction
* Program Definition
* Saving and Loading a Trained Model
* Interacting with a Trained Model

Student Materials: Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work

# Machine Learning in Java

**Course Snapshot**

* **Course:** Machine Learning in Java
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning in Java skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to know the Leverage the power of Java and its associated machine learning libraries to build powerful predictive models
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

As the amount of data in the world continues to grow at an almost incomprehensible rate, being able to understand and process data is becoming a key differentiator for competitive organizations. Machine learning applications are everywhere, from self-driving cars, spam detection, document search, and trading strategies, to speech recognition. This makes machine learning well-suited to the present-day era of big data and Data Science. The main challenge is how to transform data into actionable knowledge. Machine Learning in Java will provide you with the techniques and tools you need. You will start by learning how to apply machine learning methods to a variety of common tasks including classification, prediction, forecasting, market basket analysis, and clustering. The code in this course works for JDK 8 and above, the code is tested on JDK 11. Moving on, you will discover how to detect anomalies and fraud, and ways to perform activity recognition, image recognition, and text analysis. By the end of the course, you will have explored related web resources and technologies that will help you take your learning to the next level. By applying the most effective machine learning methods to real-world problems, you will gain hands-on experience that will transform the way you think about data.

Working in a hands-on learning environment, led by our Machine Learning in Java expert instructor, students will learn about and explore:

* Solve predictive modeling problems using the most popular machine learning Java libraries
* Explore data processing, machine learning, and NLP concepts using JavaML, WEKA, MALLET libraries
* Practical examples, tips, and tricks to help you understand applied machine learning in Java

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Discover key Java machine learning libraries
* Implement concepts such as classification, regression, and clustering
* Develop a customer retention strategy by predicting likely churn candidates
* Build a scalable recommendation engine with Apache Mahout
* Apply machine learning to fraud, anomaly, and outlier detection
* Experiment with deep learning concepts and algorithms
* Write your own activity recognition model for eHealth applications

**Audience & Pre-Requisites**

This course is geared for attendees wants to know the Leverage the power of Java and its associated machine learning libraries to build powerful predictive models.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Python programming knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Applied Machine Learning Quick Start**

* Applied Machine Learning Quick Start
* Machine learning and data science
* Data and problem definition
* Data collection
* Data preprocessing
* Unsupervised learning
* Supervised learning
* Generalization and evaluation

1. **Java Libraries and Platforms for Machine Learning**

* Java Libraries and Platforms for Machine Learning
* The need for Java
* Machine learning libraries
* Building a machine learning application

1. **Basic Algorithms - Classification, Regression, and Clustering**

* Basic Algorithms - Classification, Regression, and Clustering
* Before you start
* Classification
* Regression
* Clustering

1. **Customer Relationship Prediction with Ensembles**

* Customer Relationship Prediction with Ensembles
* The customer relationship database
* Basic Naive Bayes classifier baseline
* Basic modeling
* Advanced modeling with ensembles

1. **Affinity Analysis**

* Affinity Analysis
* Market basket analysis
* Association rule learning
* The supermarket dataset
* Discover patterns
* Other applications in various areas

1. **Recommendation Engines with Apache Mahout**

* Recommendation Engines with Apache Mahout
* Basic concepts
* Getting Apache Mahout
* Building a recommendation engine
* Content-based filtering

1. **Fraud and Anomaly Detection**

* Fraud and Anomaly Detection
* Suspicious and anomalous behavior detection
* Suspicious pattern detection
* Anomalous pattern detection
* Outlier detection using ELKI
* Fraud detection in insurance claims
* Anomaly detection in website traffic

1. **Image Recognition with Deeplearning4j**

* Image Recognition with Deeplearning4j
* Introducing image recognition
* Image classification

1. **Activity Recognition with Mobile Phone Sensors**

* Activity Recognition with Mobile Phone Sensors
* Introducing activity recognition
* Collecting data from a mobile phone
* Building a classifier

1. **Text Mining with Mallet - Topic Modeling and Spam Detection**

* Text Mining with Mallet - Topic Modeling and Spam Detection
* Introducing text mining
* Installing Mallet
* Working with text data
* Topic modeling for BBC News
* Detecting email spam

Student Materials: Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work

# What's New in TensorFlow 2.0

**Course Snapshot**

* **Course:** What's New in TensorFlow 2.0
* **Duration:** 2 days
* **Skill-level**: Foundation-level TensorFlow skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants Get grips with key structural changes in TensorFlow 2.0
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

TensorFlow is an end-to-end machine learning platform for experts as well as beginners, and its new version, TensorFlow 2.0 (TF 2.0), improves its simplicity and ease of use. This course will help you understand and utilize the latest TensorFlow features. What's New in TensorFlow 2.0 starts by focusing on advanced concepts such as the new TensorFlow Keras APIs, eager execution, and efficient distribution strategies that help you to run your machine learning models on multiple GPUs and TPUs. The course then takes you through the process of building data ingestion and training pipelines, and it provides recommendations and best practices for feeding data to models created using the new tf.keras API. You will explore the process of building an inference pipeline using TF Serving and other multi-platform deployments before moving on to explore the newly released AIY, which is essentially do-it-yourself AI. This course delves into the core APIs to help you build unified convolutional and recurrent layers and use TensorBoard to visualize deep learning models using what-if analysis. By the end of the course, you'll have learned about compatibility between TF 2.0 and TF 1.x and be able to migrate to TF 2.0 smoothly.

Working in a hands-on learning environment, led by our TensorFlow expert instructor, students will learn about and explore:

* Explore TF Keras APIs and strategies to run GPUs, TPUs, and compatible APIs across the TensorFlow ecosystem
* Learn and implement best practices for building data ingestion pipelines using TF 2.0 APIs
* Migrate your existing code from TensorFlow 1.x to TensorFlow 2.0 seamlessly

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Implement tf.keras APIs in TF 2.0 to build, train, and deploy production-grade models
* Build models with Keras integration and eager execution
* Explore distribution strategies to run models on GPUs and TPUs
* Perform what-if analysis with TensorBoard across a variety of models
* Discover Vision Kit, Voice Kit, and the Edge TPU for model deployments
* Build complex input data pipelines for ingesting large training datasets

**Audience & Pre-Requisites**

This course is geared for attendees to Get grips with key structural changes in TensorFlow 2.0

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Tensorflow knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Getting Started with TensorFlow 2.0**

* Getting Started with TensorFlow 2.0
* Technical requirements
* What's new?
* TF 2.0 installation and setup
* Using TF 2.0
* Rich extensions

1. **Keras Default Integration and Eager Execution**

* Keras Default Integration and Eager Execution
* Technical requirements
* New abstractions in TF 2.0
* Diving deep into the Keras API
* Estimators
* Evaluating TensorFlow graphs

1. **Designing and Constructing Input Data Pipelines**

* Designing and Constructing Input Data Pipelines
* Technical requirements
* Designing and constructing the data pipeline
* Transforming datasets
* Feeding the created dataset to the model
* Examples of complete end-to-end data pipelines
* Best practices and the performance optimization of a data pipeline in TF 2.0
* Built-in datasets in TF 2.0

1. **Model Training and Use of TensorBoard**

* Model Training and Use of TensorBoard
* Technical requirements
* Comparing Keras and tf.keras
* Creating models using tf.keras 2.0
* Model compilation and training
* Custom training logic
* Distributed training
* TensorBoard

1. **Model Inference Pipelines - Multi-platform Deployments**

* Model Inference Pipelines - Multi-platform Deployments
* Technical requirements
* Machine learning workflow – the inference phase
* Model artifact – the SavedModel format
* Inference on backend servers
* Inference in the browser
* Inference on mobile and IoT devices

1. **AIY Projects and TensorFlow Lite**

* AIY Projects and TensorFlow Lite
* Introduction to TFLit
* Getting started with TFLite
* Running TFLite on mobile devices
* Running TFLite on low-power machines
* Comparing TFLite and TF
* AIY

1. **Migrating From TensorFlow 1.x to 2.0**

* Migrating From TensorFlow 1.x to 2.0
* Major changes in TF 2.
* Recommended techniques to employ for idiomatic TF 2.0
* Making code TF 2.0-native
* Frequently asked questions
* The future of TF 2.0

Student Materials: Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Blockchain Development for Finance Projects

**Course Snapshot**

* **Course:** Blockchain Development for Finance Projects
* **Duration:** 4 days
* **Skill-level**: Foundation-level Blockchain Development for Finance skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to implement and re-engineering banking and financial solutions and workflows using eight step-by-step projects
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Blockchain technology will continue to play an integral role in the banking and finance sector in the coming years. It will enable enterprises to build transparent and secure business processes. Experts estimate annual savings of up to 20 billion dollars from this technology. This course will help you build financial apps using blockchain, guiding you through enhancing popular products and services in the banking and finance sector. The course starts by explaining the essential concepts of blockchain, and the impact of blockchain technology on the BFSI sector. Next, you'll delve into re-designing existing banking processes and building new financial apps using blockchain. To accomplish this, you'll work through eight blockchain projects. By demonstrating the entire process, the coourse helps you understand everything from setting up the environment and building frontend portals to system integration and testing apps. You will gain hands-on experience with the Ethereum, Hyperledger Fabric, and Stellar to develop private and public decentralized apps. Finally, you'll learn how to use ancillary platforms and frameworks such as IPFS, Truffle OpenZeppelin, and MetaMask.

.

Working in a hands-on learning environment, led by our Blockchain Development for Finance expert instructor, students will learn about and explore:

* Implement various end-to-end blockchain projects and learn to enhance present-day financial solutions
* Use Ethereum, Hyperledger, and Stellar to build public and private decentralized applications
* Address complex challenges faced in the BFSI domain using different blockchain platform services

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Design and implement blockchain solutions in a BFSI organization
* Explore common architectures and implementation models for enterprise blockchain
* Design blockchain wallets for multi-purpose applications using Ethereum
* Build secure and fast decentralized trading ecosystems with Blockchain
* Implement smart contracts to build secure process workflows in Ethereum and Hyperledger Fabric
* Use the Stellar platform to build KYC and AML-compliant remittance workflows
* Map complex business workflows and automate backend processes in a blockchain architecture

**Audience & Pre-Requisites**

This course is geared for attendees with basic Blockchain Development for Finance Projects who wish to have an in-depth understanding of how to leverage distributed ledgers and smart contracts for financial use cases.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Blockchain Development for Finance Projects knoledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Section 1: Blockchain Payments and Remittances**
2. **Blockchain in Financial Services**

* Present-day banking and finance systems
* Understanding blockchain technology
* Blockchains for financial services
* How to approach implementing a blockchain solution
* Implementation strategies
* Popular distributed ledger platforms for financial applications

1. **Building a Blockchain Wallet for Fungible and Non-Fungible Assets**

* Technical requirements
* Understanding ERC20 and ERC721 smart contract standards
* Writing the smart contract code
* Migrating the smart contract code using Truffle
* Creating the token wallet frontend using ReactJS
* Running our app
* Connecting to the main Ethereum network

1. **Designing a Payment Gateway for Online Merchants**

* Technical requirements
* Defining our blockchain payment ecosystem
* Generating dynamic merchant addresses using HD wallets
* Creating an e-commerce website and payment gateway
* Creating an API for generating dynamic payment addresses
* Building the merchant HD wallet
* Running the payment ecosystem

1. **Corporate Remittances and Settlement**

* Technical requirements
* Understanding the blockchain corporate remittance application and network layout
* Setting up the Hyperledger Fabric Bankchain network
* Creating blockchain identities for the banks
* Building the corporate remittance contract
* Setting up the IPFS network
* Setting up the bank databases
* Building the bank backend servers
* Building the transaction listeners for the banks
* Creating the corporate remittance app frontend in React
* Running the corporate remittance app

1. **Enabling Cross-Border Remittances with Real-Time KYC/AML Verification**

* Technical requirements
* Designing a workflow for blockchain cross-border remittance
* Setting up a test network
* Creating user accounts
* Setting up the bank domains
* Setting up the federation servers
* Setting up the compliance server
* Setting up the bridge server
* Setting up the callbacks server
* Building the bank portal
* Running the remittance platform

1. **Section 2: Blockchain Workflows Using Smart Contracts**
2. **Building a Letter of Credit Workflow Module Using Smart Contracts**

* Technical requirements
* Understanding smart contracts and blockchain-based workflows
* Creating a USD token for accounting
* Deploying a USD token for accounting
* Creating an LC Master smart contract
* Creating an LC smart contract
* Deploying the LC Master smart contract
* Creating the LC module React app
* Running the LC module

1. **Section 3: Securing Digital Documents and Files Using Blockchain**
2. **Building a Tamper-Proof Document Storage System**

* Technical requirements
* Tamper-proof document storage using blockchain
* Setting up the Hyperledger Fabric network
* Writing and deploying the DocsApp chaincode
* Building the backend services
* Creating a React frontend for the app
* Running the tamper-proof application

1. **Section 4: Decentralized Trading Exchanges Using Blockchain**
2. **Building a Decentralized Trading Exchange**

* Technical requirements
* Decentralized trading exchanges
* Issuing the trading assets
* Orderbook smart contract
* Building the exchange app
* Running the exchange app

1. **Developing a Currency Trading Exchange for Market Making**

* Technical requirements
* Introducing the distributed currency trading exchange
* Building the private test Stellar network
* Creating the user accounts
* Creating trading currency assets
* Building the currency trading exchange
* Running the currency exchange

1. **Looking into the Future**

* Summarizing our journey
* Extending concepts to other applications
* The road ahead – some additional blockchain concepts

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning Bootcamp

**Course Snapshot**

* **Course:** Machine Learning Bootcamp
* **Duration:** 2 days
* **Skill-level**: Foundation-level Data Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to build a portfolio of business-relevant machine learning projects that hiring managers will be excited to see.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In **Machine Learning** you’ll learn the essentials of machine learning by completing a carefully designed set of real-world projects. Beginning as a novice, you’ll start with the basic concepts of ML before tackling your first challenge: creating a car price predictor using linear regression algorithms. You’ll then advance through increasingly difficult projects, developing your skills to build a churn prediction application, a flight delay calculator, an image classifier, and more. When you’re done working through these fun and informative projects, you’ll have a comprehensive machine learning skill set you can apply to practical on-the-job problems.

Working in a hands-on learning environment, led by Machine Learning expert instructor, students will learn about and explore:

* you’ll create and deploy Python-based machine learning models for a variety of increasingly challenging projects
* you’ll learn the essentials of machine learning by completing a carefully designed set of real-world projects. Beginning as a novice
* you’ll start with the basic concepts of ML before tackling your first challenge: creating a car price predictor using linear regression algorithms.
* You’ll then advance through increasingly difficult projects, developing your skills to build a churn prediction application, a flight delay calculator, an image classifier, and more.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Code fundamental ML algorithms from scratch
* Collect and clean data for training models
* Use popular Python tools, including NumPy, Pandas, Scikit-Learn, and TensorFlow
* Apply ML to complex datasets with images and text
* Deploy ML models to a production-ready environment

**Audience & Pre-Requisites**

This course is for readers want to learn the essentials of machine learning by completing a carefully designed set of real-world projects.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.
* No previous machine learning experience required

**Course Agenda / Topics**

1. **Introduction to machine learning**

* Machine learning
* Machine learning process
* Modeling and model validation

1. **Machine learning for regression**

* Car-price prediction project
* Exploratory data analysis
* Machine learning for regression
* Predicting the price

1. **Machine learning for classification**

* Churn prediction project
* Feature engineering
* Machine learning for classification

1. **Evaluation metrics for classification**

* Evaluation metrics
* Confusion table
* ROC curve and AUC score
* Parameter tuning

1. **Deploying machine learning models**

* Churn prediction model
* Model serving
* Managing dependencies
* Deployment

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with scikit-learn

**Course Snapshot**

* **Course:** Machine Learning with scikit-learn
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning with scikit-learn skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Deploy supervised and unsupervised machine learning algorithms using scikit-learn to perform classification, regression, and clustering.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Scikit-learn is a robust machine learning library for the Python programming language. It provides a set of supervised and unsupervised learning algorithms. This course is the easiest way to learn how to deploy, optimize, and evaluate all of the important machine learning algorithms that scikit-learn provides. This course teaches you how to use scikit-learn for machine learning. You will start by setting up and configuring your machine learning environment with scikit-learn. To put scikit-learn to use, you will learn how to implement various supervised and unsupervised machine learning models. You will learn classification, regression, and clustering techniques to work with different types of datasets and train your models. Finally, you will learn about an effective pipeline to help you build a machine learning project from scratch. By the end of this course, you will be confident in building your own machine learning models for accurate predictions.

Working in a hands-on learning environment, led by Machine Learning with scikit-learn expert instructor, students will learn about and explore:

* Build your first machine learning model using scikit-learn
* Train supervised and unsupervised models using popular techniques such as classification, regression and clustering
* Understand how scikit-learn can be applied to different types of machine learning problems

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn how to work with all scikit-learn's machine learning algorithms
* Install and set up scikit-learn to build your first machine learning model
* Employ Unsupervised Machine Learning Algorithms to cluster unlabelled data into groups
* Perform classification and regression machine learning
* Use an effective pipeline to build a machine learning project from scratch

**Audience & Pre-Requisites**

This course is for readers want to Deploy supervised and unsupervised machine learning algorithms using scikit-learn to perform classification, regression, and clustering.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

1. **Introducing Machine Learning with scikit-learn**

* Introducing Machine Learning with scikit-learn
* A brief introduction to machine learning
* What is scikit-learn?
* Installing scikit-learn
* Algorithms that you will learn to implement using scikit-learn

1. **Predicting Categories with K-Nearest Neighbors**

* Predicting Categories with K-Nearest Neighbors
* Technical requirements
* Preparing a dataset for machine learning with scikit-learn
* The k-NN algorithm
* Implementing the k-NN algorithm using scikit-learn
* Fine-tuning the parameters of the k-NN algorithm
* Scaling for optimized performance

1. **Predicting Categories with Logistic Regression**

* Predicting Categories with Logistic Regression
* Technical requirements
* Understanding logistic regression mathematically
* Implementing logistic regression using scikit-learn
* Fine-tuning the hyperparameters
* Scaling the data
* Interpreting the logistic regression model

1. **Predicting Categories with Naive Bayes and SVMs**

* Predicting Categories with Naive Bayes and SVMs
* Technical requirements
* The Naive Bayes algorithm
* Support vector machines

1. **Predicting Numeric Outcomes with Linear Regression**

* Predicting Numeric Outcomes with Linear Regression
* Technical requirements
* The inner mechanics of the linear regression algorithm
* Implementing linear regression in scikit-learn
* Model optimization

1. **Classification and Regression with Trees**

* Classification and Regression with Trees
* Technical requirements
* Classification trees
* Regression trees
* Ensemble classifier

1. **Clustering Data with Unsupervised Machine Learning**

* Clustering Data with Unsupervised Machine Learning
* Technical requirements
* The k-means algorithm
* Implementing the k-means algorithm in scikit-learn
* Feature engineering for optimization
* Cluster visualization
* Going from unsupervised to supervised learning

1. **Performance Evaluation Methods**

* Performance Evaluation Methods
* Technical requirements
* Why is performance evaluation critical?
* Performance evaluation for classification algorithms
* Performance evaluation for regression algorithms
* Performance evaluation for unsupervised algorithms

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Time Series Analysis with R

**Course Snapshot**

* **Course:** Time Series Analysis with R
* **Duration:** 3 days
* **Skill-level**: Foundation-level Time Series Analysis with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Build efficient forecasting models using traditional time series models and machine learning algorithms.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Time series analysis is the art of extracting meaningful insights from, and revealing patterns in, time series data using statistical and data visualization approaches. These insights and patterns can then be utilized to explore past events and forecast future values in the series. This course explores the basics of time series analysis with R and lays the foundations you need to build forecasting models. You will learn how to preprocess raw time series data and clean and manipulate data with packages such as stats, lubridate, xts, and zoo. You will analyze data and extract meaningful information from it using both descriptive statistics and rich data visualization tools in R such as the TSstudio, plotly, and ggplot2 packages. The later section of the course delves into traditional forecasting models such as time series linear regression, exponential smoothing (Holt, Holt-Winter, and more) and Auto-Regressive Integrated Moving Average (ARIMA) models with the stats and forecast packages. You'll also cover advanced time series regression models with machine learning algorithms such as Random Forest and Gradient Boosting Machine using the h2o package. By the end of this course, you will have the skills needed to explore your data, identify patterns, and build a forecasting model using various traditional and machine learning methods.

Working in a hands-on learning environment, led by Time Series Analysis with R expert instructor, students will learn about and explore:

* Perform time series analysis and forecasting using R packages such as Forecast and h2o
* Develop models and find patterns to create visualizations using the TSstudio and plotly packages
* Master statistics and implement time-series methods using examples mentioned

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Visualize time series data and derive better insights
* Explore auto-correlation and master statistical techniques
* Use time series analysis tools from the stats, TSstudio, and forecast packages
* Explore and identify seasonal and correlation patterns
* Work with different time series formats in R
* Explore time series models such as ARIMA, Holt-Winters, and more
* Evaluate high-performance forecasting solutions

**Audience & Pre-Requisites**

This course is for readers want to Build efficient forecasting models using traditional time series models and machine learning algorithms..

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

1. **Introduction to Time Series Analysis and R**

* Introduction to Time Series Analysis and R
* Technical requirements
* Time series data
* Historical background of time series analysis
* Time series analysis
* Getting started with R
* A brief introduction to R
* Working and manipulating data

1. **Working with Date and Time Objects**

* Working with Date and Time Objects
* Technical requirements
* The date and time formats
* Date and time objects in R
* Creating a date or time index
* Manipulation of date and time with the lubridate package

1. **The Time Series Object**

* The Time Series Object
* Technical requirement
* The Natural Gas Consumption dataset
* The attributes of the ts class
* Data manipulation of ts objects
* Visualizing ts and mts objects

1. **Working with zoo and xts Objects**

* Working with zoo and xts Objects
* Technical requirement
* The zoo class
* The xts class
* Manipulating the zoo and xts objects
* Plotting zoo and xts objects
* xts, zoo, or ts – which one to use?

1. **Decomposition of Time Series Data**

* Decomposition of Time Series Data
* Technical requirement
* The moving average function
* The time series components
* The additive versus the multiplicative model
* The decomposition of time series
* Seasonal adjustment

1. **Seasonality Analysis**

* Seasonality Analysis
* Technical requirement
* Seasonality types
* Seasonal analysis with descriptive statistics
* Structural tools for seasonal analysis

1. **Correlation Analysis**

* Correlation Analysis
* Technical requirement
* Correlation between two variables
* Lags analysis
* The autocorrelation function
* The partial autocorrelation function
* Lag plots
* Causality analysis

1. **Forecasting Strategies**

* Forecasting Strategies
* Technical requirement
* The forecasting workflow
* Training approaches
* Finalizing the forecast
* Handling forecast uncertainty

1. **Forecasting with Linear Regression**

* Forecasting with Linear Regression
* Technical requirement
* The linear regression
* Forecasting with linear regression
* Forecasting a series with multiseasonality components – a case study

1. **Forecasting with Exponential Smoothing Models**

* Forecasting with Exponential Smoothing Models
* Technical requirement
* Forecasting with moving average models
* Forecasting with exponential smoothing

1. **Forecasting with ARIMA Models**

* Forecasting with ARIMA Models
* Technical requirement
* The stationary process
* The AR process
* The moving average process
* The ARMA model
* Forecasting AR, MA, and ARMA models
* The ARIMA model
* The seasonal ARIMA model
* The auto.arima function
* Linear regression with ARIMA errors

1. **Forecasting with Machine Learning Models**

* Forecasting with Machine Learning Models
* Technical requirement
* Why and when should we use machine learning?
* Why h2o?
* Forecasting monthly vehicle sales in the US – a case study

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Unsupervised Learning with Python

**Course Snapshot**

* **Course:** Unsupervised Learning with Python
* **Duration:** 2 days
* **Skill-level**: Foundation-level Time Series Analysis with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Design clever algorithms that can uncover interesting structures and hidden relationships in unstructured, unlabeled data.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Unsupervised learning is a useful and practical solution in situations where labeled data is not available. Applied Unsupervised Learning with Python guides you on the best practices for using unsupervised learning techniques in tandem with Python libraries and extracting meaningful information from unstructured data. The course begins by explaining how basic clustering works to find similar data points in a set. Once you are well versed with the k-means algorithm and how it operates, you’ll learn what dimensionality reduction is and where to apply it. As you progress, you’ll learn various neural network techniques and how they can improve your model. While studying the applications of unsupervised learning, you will also understand how to mine topics that are trending on Twitter and Facebook and build a news recommendation engine for users. You will complete the course by challenging yourself through various interesting activities such as performing a Market Basket Analysis and identifying relationships between different merchandises. By the end of this course, you will have the skills you need to confidently build your own models using Python.

Working in a hands-on learning environment, led by Unsupervised Learning with Python expert instructor, students will learn about and explore:

Learn how to select the most suitable Python library to solve your problem

Compare k-Nearest Neighbor (k-NN) and non-parametric methods and decide when to use them

Delve into the applications of neural networks using real-world datasets

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

Understand the basics and importance of clustering

Build k-means, hierarchical, and DBSCAN clustering algorithms from scratch with built-in packages

Explore dimensionality reduction and its applications

Use scikit-learn (sklearn) to implement and analyse principal component analysis (PCA)on the Iris dataset

Employ Keras to build autoencoder models for the CIFAR-10 dataset

Apply the Apriori algorithm with machine learning extensions (Mlxtend) to study transaction data

**Audience & Pre-Requisites**

This course is for readers want to Design clever algorithms that can uncover interesting structures and hidden relationships in unstructured, unlabeled data.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT and Python Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

1. **Introduction to Clustering**

Introduction to Clustering

Introduction

Unsupervised Learning versus Supervised Learning

Clustering

Introduction to k-means Clustering

Activity 1: Implementing k-means Clustering

1. **Hierarchical Clustering**

Hierarchical Clustering

Introduction

Clustering Refresher

The Organization of Hierarchy

Introduction to Hierarchical Clustering

Linkage

Agglomerative versus Divisive Clustering

k-means versus Hierarchical Clustering

1. **Neighborhood Approaches and DBSCAN**

Neighborhood Approaches and DBSCAN

Introduction

Introduction to DBSCAN

DBSCAN Versus k-means and Hierarchical Clustering

1. **Dimension Reduction and PCA**

Dimension Reduction and PCA

Introduction

Overview of Dimensionality Reduction Techniques

PCA

1. **Autoencoders**

Autoencoders

Introduction

Fundamentals of Artificial Neural Networks

Autoencoders

1. **t-Distributed Stochastic Neighbor Embedding (t-SNE)**

t-Distributed Stochastic Neighbor Embedding (t-SNE)

Introduction

Stochastic Neighbor Embedding (SNE)

t-Distributed SNE

Interpreting t-SNE Plots

1. **Topic Modeling**

Topic Modeling

Introduction

Cleaning Text Data

Latent Dirichlet Allocation

Non-Negative Matrix Factorization

1. **Market Basket Analysis**

Market Basket Analysis

Introduction

Market Basket Analysis

Characteristics of Transaction Data

Apriori Algorithm

Association Rules

1. **Hotspot Analysis**

Hotspot Analysis

Introduction

Kernel Density Estimation

Hotspot Analysis

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Applied Unsupervised Learning with Python

**Course Snapshot**

* **Course:** Applied Unsupervised Learning with Python
* **Duration:** 2 days
* **Skill-level**: Foundation-level Time Series Analysis with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Discover the skill-sets required to implement various approaches to Machine Learning with Python.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Unsupervised learning is about making use of raw, untagged data and applying learning algorithms to it to help a machine predict its outcome. With this course, you will explore the concept of unsupervised learning to cluster large sets of data and analyze them repeatedly until the desired outcome is found using Python. This course starts with the key differences between supervised, unsupervised, and semi-supervised learning. You will be introduced to the best-used libraries and frameworks from the Python ecosystem and address unsupervised learning in both the machine learning and deep learning domains. You will explore various algorithms, techniques that are used to implement unsupervised learning in real-world use cases. You will learn a variety of unsupervised learning approaches, including randomized optimization, clustering, feature selection and transformation, and information theory. You will get hands-on experience with how neural networks can be employed in unsupervised scenarios. You will also explore the steps involved in building and training a GAN in order to process images. By the end of this course, you will have learned the art of unsupervised learning for different real-world challenges.

Working in a hands-on learning environment, led by Unsupervised Learning with Python expert instructor, students will learn about and explore:

* Explore unsupervised learning with clustering, autoencoders, restricted Boltzmann machines, and more
* Build your own neural network models using modern Python libraries
* Practical examples show you how to implement different machine learning and deep learning techniques

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use cluster algorithms to identify and optimize natural groups of data
* Explore advanced non-linear and hierarchical clustering in action
* Soft label assignments for fuzzy c-means and Gaussian mixture models
* Detect anomalies through density estimation
* Perform principal component analysis using neural network models
* Create unsupervised models using GANs

**Audience & Pre-Requisites**

This course is for readers want to Discover the skill-sets required to implement various approaches to Machine Learning with Python.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT and Python Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

1. **Getting Started with Unsupervised Learning**

* Getting Started with Unsupervised Learning
* Technical requirements
* Why do we need machine learning?
* Types of machine learning algorithm
* Why Python for data science and machine learning?

1. **Clustering Fundamentals**

* Clustering Fundamentals
* Technical requirements
* Introduction to clustering
* K-means
* Analysis of the Breast Cancer Wisconsin dataset
* Evaluation metrics
* K-Nearest Neighbors
* Vector Quantization

1. **Advanced Clustering**

* Advanced Clustering
* Technical requirements
* Spectral clustering
* Mean shift
* DBSCAN
* K-medoids
* Online clustering

1. **Hierarchical Clustering in Action**

* Hierarchical Clustering in Action
* Technical requirements
* Cluster hierarchies
* Agglomerative clustering
* Analyzing a dendrogram
* Cophenetic correlation as a performance metric
* Agglomerative clustering on the Water Treatment Plant dataset
* Connectivity constraints

1. **Soft Clustering and Gaussian Mixture Models**

* Soft Clustering and Gaussian Mixture Models
* Technical requirements
* Soft clustering
* Fuzzy c-means
* Gaussian mixture

1. **Anomaly Detection**

* Anomaly Detection
* Technical requirements
* Probability density functions
* Histograms
* Kernel density estimation (KDE)
* Anomaly detection
* One-class support vector machines
* Anomaly detection with Isolation Forests

1. **Dimensionality Reduction and Component Analysis**

* Dimensionality Reduction and Component Analysis
* Technical requirements
* Principal Component Analysis (PCA)
* Independent Component Analysis
* Topic modeling with Latent Dirichlet Allocation

1. **Unsupervised Neural Network Models**

* Unsupervised Neural Network Models
* Technical requirements
* Autoencoders
* Hebbian-based principal component analysis
* Unsupervised deep belief networks

1. **Generative Adversarial Networks and SOMs**

* Generative Adversarial Networks and SOMs
* Technical requirements
* Generative adversarial networks
* Self-organizing maps

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning.

**Course Snapshot**

* **Course:** Machine Learning
* **Duration:** 3 days
* **Skill-level**: Foundation-level machine learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending how to apply ML to your projects using only standard Python code and high school-level math. No specialist knowledge is required to tackle the hands-on exercises using readily-available machine learning tools!
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In **Machine Learning**, expert machine learning engineer Luis Serrano introduces the most valuable ML techniques and teaches you how to make them work for you. You’ll only need high school math to dive into popular approaches and algorithms. Practical examples illustrate each new concept to ensure you’re grokking as you go. You’ll build models for spam detection, language analysis, and image recognition as you lock in each carefully-selected skill. Packed with easy-to-follow Python-based exercises and mini-projects, this course sets you on the path to becoming a machine learning expert. When you’re done, you’ll have an intuitive understanding of the right approach for any machine learning task or project.

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* how to apply ML to your projects using only standard Python code and high school-level math.
* No specialist knowledge is required to tackle the hands-on exercises using readily-available machine learning tools!

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Different types of machine learning, including supervised and unsupervised learning
* Algorithms for simplifying, classifying, and splitting data
* Machine learning packages and tools
* Hands-on exercises with fully-explained Python code samples

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to learn and use basic machine learning algorithms and concepts.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* No machine learning experience or advanced math skills necessary.
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **What is machine learning?**

* Why this course?
* Is machine learning hard?
* But what exactly is machine learning?
* Humans use the remember-formulate-predict framework to make decisions (and so can machines!)
* What is this course about?

1. **Types of machine learning**

* What is the difference between labelled and unlabeled data?
* What is supervised learning?
* What is unsupervised learning?
* What is reinforcement learning?

1. **Drawing a line close to our points: Linear regression**

* The problem: We need to predict the price of a house
* The solution: Building a regression model for housing prices
* How to get the computer to draw this line: the linear regression algorithm
* Applications of linear regression

1. **Using lines to split our points: The perceptron algorithm**

* The problem: We are in an alien planet, and we don’t know their language!
* How do we determine if a classifier is good or bad? The error functions
* How to find a good classifier? The perceptron algorithm
* Coding the perceptron algorithm
* Applications
* Some drawbacks of the perceptron algorithm, which will be addressed very soon!

1. **A continuous approach to splitting points: Logistic regression**

* Logistic Regression (or continuous perceptron)
* Reducing the log loss error: The logistic regression trick
* Classifying into multiple classes - The SoftMax function

1. **Using probability to its maximum: The naive Bayes algorithm**

* Sick or healthy? A story with Bayes Theorem
* Use-case: Spam detection model
* Building a spam detection model with real data

1. **Splitting data by asking questions: Decision trees**

* The problem: We need to recommend apps to users according to what they are likely to download
* The solution: Building an app recommendation system
* Building the tree: How to pick the right feature to split
* Back to recommending apps: Building our decision tree using Gini index
* Beyond questions like yes/no
* Coding a decision tree with sk learn
* A slightly larger example: Spam detection again!
* Applications

1. **Combining models to maximize: Ensemble learning**

* With a little help from our friends
* Why an ensemble of learners? Why not just one really good learner?
* Bagging - Joining some classifiers together to build a stronger classifier
* Boosting - Joining some classifiers together in a smarter way to get a stronger classifier
* Applications of ensemble metho

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning Overview.

**Course Snapshot**

* **Course:** Machine Learning Overview
* **Duration:** 3 days
* **Skill-level**: Foundation-level machine learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to get guide to developing, training, and optimizing your machine learning models
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning makes it possible to learn about the unknowns and gain hidden insights into your datasets by mastering many tools and techniques. This course guides you to do just that in a very compact manner. After giving a quick overview of what machine learning is all about, Machine Learning Quick Reference jumps right into its core algorithms and demonstrates how they can be applied to real-world scenarios. From model evaluation to optimizing their performance, this course will introduce you to the best practices in machine learning. Furthermore, you will also look at the more advanced aspects such as training neural networks and work with different kinds of data, such as text, time-series, and sequential data. Advanced methods and techniques such as causal inference, deep Gaussian processes, and more are also covered. By the end of this course, you will be able to train fast, accurate machine learning models at your fingertips, which you can easily use as a point of reference

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Your guide to learning efficient machine learning processes from scratch
* Explore expert techniques and hacks for a variety of machine learning concepts
* Write effective code in R, Python, Scala, and Spark to solve all your machine learning problems

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get a quick rundown of model selection, statistical modeling, and cross-validation
* Choose the best machine learning algorithm to solve your problem
* Explore kernel learning, neural networks, and time-series analysis
* Train deep learning models and optimize them for maximum performance
* Briefly cover Bayesian techniques and sentiment analysis in your NLP solution
* Implement probabilistic graphical models and causal inferences
* Measure and optimize the performance of your machine learning models

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to get guide to developing, training, and optimizing your machine learning models

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* No machine learning experience or advanced math skills necessary.
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Quantifying Learning Algorithms**

* Quantifying Learning Algorithms
* Statistical models
* Learning curve
* Curve fitting
* Statistical modeling – the two cultures of Leo Breiman
* Training data development data – test data
* Bias-variance trade off
* Regularization
* Cross-validation and model selection
* Model selection using cross-validation
* 0.632 rule in bootstrapping
* Model evaluation
* Receiver operating characteristic curve
* H-measure
* Dimensionality reduction

1. **Evaluating Kernel Learning**

* Evaluating Kernel Learning
* Introduction to vectors
* Linear separability
* Hyperplanes
* SVM
* Kernel trick
* Kernel types
* SVM example and parameter optimization through grid search

1. **Performance in Ensemble Learning**

* Performance in Ensemble Learning
* What is ensemble learning?
* Bagging
* Decision tree
* Random forest algorithm
* Boosting

1. **Training Neural Networks**

* Training Neural Networks
* Neural networks
* Network initialization
* Overfitting
* Prevention of overfitting in NNs
* Vanishing gradient
* Recurrent neural networks

1. **Time Series Analysis**

* Time Series Analysis
* Introduction to time series analysis
* White noise
* Random walk
* Autoregression
* Autocorrelation
* Stationarity
* AR model
* Moving average model
* Autoregressive integrated moving average
* Optimization of parameters
* Anomaly detection

1. **Natural Language** **Processing**

* Natural Language Processing
* Text corpus
* TF-IDF
* Sentiment analysis
* Topic modeling
* The Bayes theorem

1. **Temporal and Sequential Pattern Discovery**

* Temporal and Sequential Pattern Discovery
* Association rules
* Apriori algorithm
* Frequent pattern growth

1. **Probabilistic Graphical Models**

* Probabilistic Graphical Models
* Key concepts
* Bayes rule
* Bayes network

1. **Selected Topics in Deep Learning**

* Selected Topics in Deep Learning
* Deep neural networks
* Backward propagation
* Forward propagation equation
* Backward propagation equation
* Parameters and hyperparameters
* Bias initialization
* Generative adversarial networks
* Hinton's Capsule network

1. **Causal Inference**

* Causal Inference
* Granger causality
* F-test
* Graphical causal models

1. **Advanced Methods**

* Advanced Methods
* Introduction
* Kernel PCA
* Independent component analysis
* Compressed sensing
* Self-organizing maps
* Bayesian multiple imputatio

**Student Materials:** Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Applied Supervised Learning with R.

**Course Snapshot**

* **Course:** Applied Supervised Learning with R
* **Duration:** 2 days
* **Skill-level**: Foundation-level Applied Supervised Learning with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Learn the ropes of supervised machine learning with R by studying popular real-world use cases, and understand how it drives object detection in driverless cars, customer churn, and loan default prediction
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

R provides excellent visualization features that are essential for exploring data before using it in automated learning. Applied Supervised Learning with R helps you cover the complete process of employing R to develop applications using supervised machine learning algorithms for your business needs. The course starts by helping you develop your analytical thinking to create a problem statement using business inputs and domain research. You will then learn different evaluation metrics that compare various algorithms, and later progress to using these metrics to select the best algorithm for your problem. After finalizing the algorithm you want to use, you will study the hyperparameter optimization technique to fine-tune your set of optimal parameters. The course demonstrates how you can add different regularization terms to avoid overfitting your model. By the end of this course, you will have gained the advanced skills you need for modeling a supervised machine learning algorithm that precisely fulfills your business needs.

Working in a hands-on learning environment, led by our Applied Supervised Learning with R expert instructor, students will learn about and explore:

* Study supervised learning algorithms by using real-world datasets
* Fine-tune optimal parameters with hyperparameter optimization
* Select the best algorithm using the model evaluation framework

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Develop analytical thinking to precisely identify a business problem
* Wrangle data with dplyr, tidyr, and reshape2
* Visualize data with ggplot2
* Validate your supervised machine learning model using k-fold
* Optimize hyperparameters with grid and random search, and Bayesian optimization
* Deploy your model on Amazon Web Services (AWS) Lambda with plumber
* Improve your model’s performance with feature selection and dimensionality reduction

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Learn the ropes of supervised machine learning with R by studying popular real-world use cases, and understand how it drives object detection in driverless cars, customer churn, and loan default prediction

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* No machine learning experience or advanced math skills necessary.
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **R for Advanced Analytics**

* R for Advanced Analytics
* Introduction
* Working with Real-World Datasets
* Reading Data from Various Data Formats
* Write R Markdown Files for Code Reproducibility
* Data Structures in R
* DataFrame
* Data Processing and Transformation
* The Apply Family of Functions
* Useful Packages
* Data Visualization
* Line Charts
* Histogram
* Boxplot

1. **Exploratory Analysis of Data**

* Exploratory Analysis of Data
* Introduction
* Defining the Problem Statement
* Understanding the Science Behind EDA
* Exploratory Data Analysis
* Univariate Analysis
* Exploring Categorical Features
* Bivariate Analysis
* Studying the Relationship between Two Numeric Variables
* Studying the Relationship between a Categorical and a Numeric Variable
* Studying the Relationship Between Two Categorical Variables
* Multivariate Analysis
* Validating Insights Using Statistical Tests
* Categorical Dependent and Numeric/Continuous Independent Variables
* Categorical Dependent and Categorical Independent Variables

1. **Introduction to Supervised Learning**

* Introduction to Supervised Learning
* Introduction
* Summary of the Beijing PM2.5 Dataset
* Regression and Classification Problems
* Machine Learning Workflow
* Regression
* Exploratory Data Analysis (EDA)
* Classification
* Evaluation Metrics

1. **Regression**

* Regression
* Introduction
* Linear Regression
* Model Diagnostics
* Residual versus Fitted Plot
* Normal Q-Q Plot
* Scale-Location Plot
* Residual versus Leverage
* Improving the Model
* Quantile Regression
* Polynomial Regression
* Ridge Regression
* LASSO Regression
* Elastic Net Regression
* Poisson Regression
* Cox Proportional-Hazards Regression Model
* NCCTG Lung Cancer Data

1. **Classification**

* Classification
* Introduction
* Getting Started with the Use Case
* Classification Techniques for Supervised Learning
* Logistic Regression
* How Does Logistic Regression Work?
* Evaluating Classification Models
* What Metric Should You Choose?
* Evaluating Logistic Regression
* Decision Trees
* XGBoost
* Deep Neural Networks
* Choosing the Right Model for Your Use Case

1. **Feature Selection and Dimensionality Reduction**

* Feature Selection and Dimensionality Reduction
* Introduction
* Feature Engineering
* One-Hot Encoding
* Log Transformation
* Feature Selection
* Highly Correlated Variables
* Feature Reduction
* Variable Clustering
* Linear Discriminant Analysis for Feature Reduction

1. **Model Improvements**

* Model Improvements
* Introduction
* Bias-Variance Trade-off
* Underfitting and Overfitting
* Defining a Sample Use Case
* Cross-Validation
* Holdout Approach/Validation
* K-Fold Cross-Validation
* Hold-One-Out Validation
* Hyperparameter Optimization
* Grid Search Optimization
* Random Search Optimization
* Bayesian Optimization

1. **Model Deployment**

* Model Deployment
* Introduction
* What is an API?
* Introduction to plumber
* A Brief History of the Pre-Docker Era
* Docker
* Amazon Web Services
* Introducing AWS SageMaker
* What is Amazon Lambda?
* What is Amazon API Gateway?
* Building Serverless ML Applications
* Deleting All Cloud Resources to Stop Billing

1. **Capstone Project - Based on Research Papers**

* Capstone Project - Based on Research Papers
* Introduction
* Exploring Research Work
* The mlr Package
* Problem Design from the Research Paper
* Features in Scene Dataset
* Implementing Multilabel Classifier Using the mlr and OpenML Packages
* Constructing a Learner
* Predictions

Student Materials: Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with TensorFlow.js.

**Course Snapshot**

* **Course:** Machine Learning with TensorFlow.js
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning with TensorFlow.js skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending Get the browser-based JavaScript library for training and deploying machine learning models effectively
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

TensorFlow.js is a framework that enables you to create performant machine learning (ML) applications that run smoothly in a web browser. With this course, you will learn how to use TensorFlow.js to implement various ML models through an example-based approach. Starting with the basics, you'll understand how ML models can be built on the web. Moving on, you will get to grips with the TensorFlow.js ecosystem to develop applications more efficiently. The course will then guide you through implementing ML techniques and algorithms such as regression, clustering, fast Fourier transform (FFT), and dimensionality reduction. You will later cover the Bellman equation to solve Markov decision process (MDP) problems and understand how it is related to reinforcement learning. Finally, you will explore techniques for deploying ML-based web applications and training models with TensorFlow Core. Throughout this ML course, you'll discover useful tips and tricks that will build on your knowledge. By the end of this course, you will be equipped with the skills you need to create your own web-based ML applications and fine-tune models to achieve high performance.

Working in a hands-on learning environment, led by our Machine Learning with TensorFlow.js expert instructor, students will learn about and explore:

* Build, train and run machine learning models in the browser using TensorFlow.js
* Create smart web applications from scratch with the help of useful examples
* Use flexible and intuitive APIs from TensorFlow.js to understand how machine learning algorithms function

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use the t-SNE algorithm in TensorFlow.js to reduce dimensions in an input dataset
* Deploy tfjs-converter to convert Keras models and load them into TensorFlow.js
* Apply the Bellman equation to solve MDP problems
* Use the k-means algorithm in TensorFlow.js to visualize prediction results
* Create tf.js packages with Parcel, Webpack, and Rollup to deploy web apps
* Implement tf.js backend frameworks to tune and accelerate app performance

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to get the browser-based JavaScript library for training and deploying machine learning models effectively.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Section 1: The Rationale of Machine Learning and the Usage of TensorFlow.js**
2. **Machine Learning for the Web**

Machine Learning for the Web

Technical requirement

Why machine learning on the web

Operation graphs

What is TensorFlow.js?

Installing TensorFlow.js

The low-level API

The Layers API

1. **Importing Pretrained Models into TensorFlow.js**

* Importing Pretrained Models into TensorFlow.js
* Technical requirements
* The portable model format
* Exporting a model from TensorFlow
* Converting models using tfjs-converter
* Loading the model into TensorFlow.js

1. **TensorFlow.js Ecosystem**

* TensorFlow.js Ecosystem
* Technical requirements
* Why high-level libraries?
* Using existing models
* Loading the data from various kinds of storage
* Pose detection with ML5.js
* Drawing cats with Magenta.js
* XOR classification with machinelearn.js

**Section 2: Real-World Applications of TensorFlow.js**

1. **Polynomial Regression**

* Polynomial Regression
* Technical requirements
* What is polynomial regression?
* Two-dimensional curve fitting

1. **Classification with Logistic Regression**

* Classification with Logistic Regression
* Technical requirements
* Background of binary classification
* What is logistic regression?
* Classifying two-dimensional clusters

1. **Unsupervised Learning**

* Unsupervised Learning
* Technical requirements
* What is unsupervised learning?
* Learning how K-means works
* Generalizing K-means with the EM algorithm
* Clustering two groups in a 2D space

1. **Sequential Data Analysis**

* Sequential Data Analysis
* Technical requirements
* What is Fourier transformation?
* Cosine curve decomposition

1. **Dimensionality Reduction**

* Dimensionality Reduction
* Technical requirements
* Why dimensionality reduction?
* Understanding principal component analysis
* Projecting 3D points into a 2D space with PCA
* Word embedding

1. **Solving the Markov Decision Process**

* Solving the Markov Decision Process
* Technical requirements
* Reinforcement learning
* Solving the four-states environment

1. **Section 3: Productionizing Machine Learning Applications with TensorFlow.js**
2. **Deploying Machine Learning Applications**

* Deploying Machine Learning Applications
* Technical requirements
* The ecosystem around the JavaScript platform
* Module bundler
* Deploying modules with GitHub Pages

1. **Tuning Applications to Achieve High Performance**

* Tuning Applications to Achieve High Performance
* Technical requirement
* The backend API of TensorFlow.js
* Tensor management
* Asynchronous data access
* Profiling
* Model visualization

1. **Work Around TensorFlow.js**

* Future Work Around TensorFlow.js
* Technical requirements
* Experimental backend implementations
* AutoML edge h

Student Materials: Each student will receive a Student Guide with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work

# Python Machine Learning.

**Course Snapshot**

* **Course:** Python Machine Learning
* **Duration:** 5 days
* **Skill-level**: Foundation-level Python Machine Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Apply machine learning with a solid foundation in theory. Revised and expanded for TensorFlow 2, GANs, and reinforcement learning.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Python Machine Learning, This Edition is a comprehensive guide to machine learning and deep learning with Python. It acts as both a step-by-step tutorial, and a reference you'll keep coming back to as you build your machine learning systems. Packed with clear explanations, visualizations, and working examples, the course covers all the essential machine learning techniques in depth. While some course teach you only to follow instructions, with this machine learning course, Raschka and Mirjalili teach the principles behind machine learning, allowing you to build models and applications for yourself. Updated for TensorFlow 2.0, this new edition introduces readers to its new Keras API features, as well as the latest additions to scikit-learn. It's also expanded to cover cutting-edge reinforcement learning techniques based on deep learning, as well as an introduction to GANs. Finally, this course also explores a subfield of natural language processing (NLP) called sentiment analysis, helping you learn how to use machine learning algorithms to classify documents. This course is your companion to machine learning with Python, whether you're a Python developer new to machine learning or want to deepen your knowledge of the latest developments.

Working in a hands-on learning environment, led by our Python Machine Learning expert instructor, students will learn about and explore:

* Widely acclaimed Python machine learning course
* Clear and intuitive explanations take you deep into the theory and practice of Python machine learning
* Fully updated and expanded to cover TensorFlow 2, Generative Adversarial Network models, reinforcement learning, and best practices

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Master the frameworks, models, and techniques that enable machines to 'learn' from data
* Use scikit-learn for machine learning and TensorFlow for deep learning
* Apply machine learning to image classification, sentiment analysis, intelligent web applications, and more
* Build and train neural networks, GANs, and other models
* Discover best practices for evaluating and tuning models
* Predict continuous target outcomes using regression analysis
* Dig deeper into textual and social media data using sentiment analysis

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Apply machine learning with a solid foundation in theory. Revised and expanded for TensorFlow 2, GANs, and reinforcement learning.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Giving Computers the Ability to Learn from Data**

* Giving Computers the Ability to Learn from Data
* Building intelligent machines to transform data into knowledge
* The three different types of machine learning
* Introduction to the basic terminology and notations
* A roadmap for building machine learning systems
* Using Python for machine learning

1. **Training Simple Machine Learning Algorithms for Classification**

* Training Simple Machine Learning Algorithms for Classification
* Artificial neurons – a brief glimpse into the early history of machine learning
* Implementing a perceptron learning algorithm in Python
* Adaptive linear neurons and the convergence of learning

1. **A Tour of Machine Learning Classifiers Using scikit-learn**

* A Tour of Machine Learning Classifiers Using scikit-learn
* Choosing a classification algorithm
* First steps with scikit-learn – training a perceptron
* Modeling class probabilities via logistic regression
* Maximum margin classification with support vector machines
* Solving nonlinear problems using a kernel SVM
* Decision tree learning
* K-nearest neighbors – a lazy learning algorithm

1. **Building Good Training Datasets – Data Preprocessing**

* Building Good Training Datasets – Data Preprocessing
* Dealing with missing data
* Handling categorical data
* Partitioning a dataset into separate training and test datasets
* Bringing features onto the same scale
* Selecting meaningful features
* Assessing feature importance with random forests

1. **Compressing Data via Dimensionality Reduction**

* Compressing Data via Dimensionality Reduction
* Unsupervised dimensionality reduction via principal component analysis
* Supervised data compression via linear discriminant analysis
* Using kernel principal component analysis for nonlinear mappings

1. **Learning Best Practices for Model Evaluation and** **Hyperparameter Tuning**

* Learning Best Practices for Model Evaluation and Hyperparameter Tuning
* Streamlining workflows with pipelines
* Using k-fold cross-validation to assess model performance
* Debugging algorithms with learning and validation curves
* Fine-tuning machine learning models via grid search
* Looking at different performance evaluation metrics

1. **Combining Different Models for Ensemble Learning**

* Combining Different Models for Ensemble Learning
* Learning with ensembles
* Combining classifiers via majority vote
* Bagging – building an ensemble of classifiers from bootstrap samples
* Leveraging weak learners via adaptive boosting

1. **Applying Machine Learning to Sentiment Analysis**

* Applying Machine Learning to Sentiment Analysis
* Preparing the IMDb movie review data for text processing
* Introducing the bag-of-words model
* Training a logistic regression model for document classification
* Working with bigger data – online algorithms and out-of-core learning
* Topic modeling with Latent Dirichlet Allocation

1. **Embedding a Machine Learning Model into a Web Application**

* Embedding a Machine Learning Model into a Web Application
* Serializing fitted scikit-learn estimators
* Setting up an SQLite database for data storage
* Developing a web application with Flask
* Turning the movie review classifier into a web application
* Deploying the web application to a public server

1. **Predicting Continuous Target Variables with Regression Analysis**

* Predicting Continuous Target Variables with Regression Analysis
* Introducing linear regression
* Exploring the Housing dataset
* Implementing an ordinary least squares linear regression model
* Fitting a robust regression model using RANSAC
* Evaluating the performance of linear regression models
* Using regularized methods for regression
* Turning a linear regression model into a curve – polynomial regression
* Dealing with nonlinear relationships using random forests

1. **Working with Unlabeled Data – Clustering Analysis**

* Working with Unlabeled Data – Clustering Analysis
* Grouping objects by similarity using k-means
* Organizing clusters as a hierarchical tree
* Locating regions of high density via DBSCAN

1. **Implementing a Multilayer Artificial Neural Network from Scratch**

* Implementing a Multilayer Artificial Neural Network from Scratch
* Modeling complex functions with artificial neural networks
* Classifying handwritten digits
* Training an artificial neural network
* About the convergence in neural networks
* A few last words about the neural network implementation

1. **Parallelizing Neural Network Training with TensorFlow**

* Parallelizing Neural Network Training with TensorFlow
* TensorFlow and training performance
* First steps with TensorFlow
* Building input pipelines using tf.data – the TensorFlow Dataset API
* Building an NN model in TensorFlow
* Choosing activation functions for multilayer neural networks

1. **Going Deeper – The Mechanics** **of TensorFlow**

* Going Deeper – The Mechanics of TensorFlow
* The key features of TensorFlow
* TensorFlow's computation graphs: migrating to TensorFlow v2
* TensorFlow Variable objects for storing and updating model parameters
* Computing gradients via automatic differentiation and GradientTape
* Simplifying implementations of common architectures via the Keras API
* TensorFlow Estimators

1. **Classifying Images with Deep Convolutional Neural Networks**

* Classifying Images with Deep Convolutional Neural Networks
* The building blocks of CNNs
* Putting everything together – implementing a CNN
* Implementing a deep CNN using TensorFlow
* Gender classification from face images using a CNN

1. **Modeling Sequential Data Using Recurrent Neural Networks**

* Modeling Sequential Data Using Recurrent Neural Networks
* Introducing sequential data
* RNNs for modeling sequences
* Implementing RNNs for sequence modeling in TensorFlow
* Understanding language with the Transformer model

1. **Generative Adversarial Networks for Synthesizing New Data**

* Generative Adversarial Networks for Synthesizing New Data
* Introducing generative adversarial networks
* Implementing a GAN from scratch
* Improving the quality of synthesized images using a convolutional and Wasserstein GAN
* Other GAN applications

1. **Reinforcement Learning for Decision Making in Complex Environments**

* Reinforcement Learning for Decision Making in Complex Environments
* Introduction – learning from experience
* The theoretical foundations of RL
* Reinforcement learning algorithms
* Implementing our first RL algorithm

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning Algorithms.

**Course Snapshot**

* **Course:** Machine Learning Algorithms
* **Duration:** 6 days
* **Skill-level**: Foundation-level Machine Learning Algorithms skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to explore and master the most important algorithms for solving complex machine learning problems
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine Learning Algorithms helps you harness the real power of machine learning algorithms to implement smarter ways of meeting today's overwhelming data needs. This newly updated and revised guide will help you master algorithms used widely in semi-supervised learning, reinforcement learning, supervised learning, and unsupervised learning domains. You will use all the modern libraries from the Python ecosystem – including NumPy and Kera’s – to extract features from varied complexities of data. Ranging from Bayesian models to the Markov chain Monte Carlo algorithm to Hidden Markov models, this machine learning course teaches you how to extract features from your dataset, perform complex dimensionality reduction, and train supervised and semi-supervised models by making use of Python-based libraries such as scikit-learn. You will also discover practical applications for complex techniques such as maximum likelihood estimation, Hebbian learning, and ensemble learning, and how to use TensorFlow 2.x to train effective deep neural networks. By the end of this course, you will be ready to implement and solve end-to-end machine learning problems and use case scenarios.

Working in a hands-on learning environment, led by our Machine Learning Algorithms expert instructor, students will learn about and explore:

* Updated to include new algorithms and techniques
* Code updated to Python 3.8 & TensorFlow 2.x
* New coverage of regression analysis, time series analysis, deep learning models, and cutting-edge applications

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the characteristics of a machine learning algorithm
* Implement algorithms from supervised, semi-supervised, unsupervised, and RL domains
* Learn how regression works in time-series analysis and risk prediction
* Create, model, and train complex probabilistic models
* Cluster high-dimensional data and evaluate model accuracy
* Discover how artificial neural networks work – train, optimize, and validate them
* Work with autoencoders, Hebbian networks, and GANs

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to learn and use basic machine learning algorithms and concepts.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Machine Learning Model Fundamentals**

* Machine Learning Model Fundamentals
* Models and data
* Characteristics of a machine learning model

1. **Loss Functions and Regularization**

* Loss Functions and Regularization
* Defining loss and cost functions
* Regularization

1. **Introduction to Semi-Supervised Learning**

* Introduction to Semi-Supervised Learning
* Semi-supervised scenario
* Generative Gaussian Mixture
* Self-Training
* Co-Training

1. **Advanced Semi-Supervised Classification**

* Advanced Semi-Supervised Classification
* Contrastive Pessimistic Likelihood Estimation
* Semi-supervised Support Vector Machines (S3VM)
* Transductive Support Vector Machines (TSVM)

1. **Graph-Based Semi-Supervised Learning**

* Graph-Based Semi-Supervised Learning
* Label propagation
* Example of label propagation
* Label spreading
* Label propagation based on Markov random walks
* Manifold learning

1. **Clustering and Unsupervised Models**

* Clustering and Unsupervised Models
* K-nearest neighbors
* K-means
* Evaluation metrics

1. **Advanced Clustering and Unsupervised Models**

* Advanced Clustering and Unsupervised Models
* Fuzzy C-means
* Spectral clustering
* DBSCAN

1. **Clustering and Unsupervised Models for Marketing**

* Clustering and Unsupervised Models for Marketing
* Biclustering
* Introduction to Market Basket Analysis with the Apriori Algorithm

1. **Generalized Linear Models and Regression**

* Generalized Linear Models and Regression
* GLMs
* Other regression techniques

1. **Introduction to Time-Series Analysis**

* Introduction to Time-Series Analysis
* Time-series
* Introduction to linear models for time-series

1. **Bayesian Networks and Hidden Markov Models**

* Bayesian Networks and Hidden Markov Models
* Conditional probabilities and Bayes' theorem
* Bayesian networks
* Hidden Markov Models

1. **The EM Algorithm**

* The EM Algorithm
* MLE and MAP Learning
* EM Algorithm
* Gaussian Mixture

1. **Component Analysis and Dimensionality Reduction**

* Component Analysis and Dimensionality Reduction
* Factor Analysis
* Principal Component Analysis
* Independent Component Analysis
* Addendum to Hidden Markov Models

1. **Hebbian Learning**

* Hebbian Learning
* Hebb's rule
* Sanger's network
* Rubner-Tavan's network
* Self-organizing maps

1. **Fundamentals of Ensemble Learning**

* Fundamentals of Ensemble Learning
* Ensemble learning fundamentals
* Random forests
* AdaBoost

1. **Advanced Boosting Algorithms**

* Advanced Boosting Algorithms
* Gradient boosting
* Ensembles of voting classifiers
* Ensemble learning as model selection

1. **17Modeling Neural Networks**

* Modeling Neural Networks
* The basic artificial neuron
* The perceptron
* Multilayer Perceptrons (MLPs)
* The back-propagation algorithm

1. **Optimizing Neural Networks**

* Optimizing Neural Networks
* Optimization algorithms
* Regularization and Dropout
* Batch normalization

1. **Deep Convolutional Networks**

* Deep Convolutional Networks
* Deep convolutional networks
* Convolutional operators
* Pooling layers
* Example of a deep convolutional network with TensorFlow and Keras

1. **Recurrent Neural Networks**

* Recurrent Neural Networks
* Recurrent networks
* Long Short-Term Memory (LSTM)
* Transfer learning

1. **Autoencoders**

* Autoencoders
* Autoencoders
* Denoising autoencoders
* Sparse autoencoders
* Variational autoencoders

1. **Introduction to Generative Adversarial Networks**

* Introduction to Generative Adversarial Networks
* Adversarial training
* Deep Convolutional GANs
* Wasserstein GAN

1. **Deep Belief Networks**

* Deep Belief Networks
* Introduction to Markov random fields
* Restricted Boltzmann Machines
* Deep Belief Networks

1. **Introduction to Reinforcement Learning**

* Introduction to Reinforcement Learning
* Fundamental concepts of RL
* Policy iteration
* Value iteration
* The TD(0) algorithm

1. **Advanced Policy Estimation Algorithms**

* Advanced Policy Estimation Algorithms
* TD(λ) algorithm
* SARSA algorithm
* Q-learning
* Direct policy search through policy gradient

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# One-shot Learning with Python

**Course Snapshot**

* **Course:** One-shot Learning with Python
* **Duration:** 2 days
* **Skill-level**: Foundation-level Data Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Get to grips with building powerful deep learning models using PyTorch and scikit-learn
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

One-shot learning has been an active field of research for scientists trying to develop a cognitive machine that mimics human learning. With this course, you'll explore key approaches to one-shot learning, such as metrics-based, model-based, and optimization-based techniques, all with the help of practical examples.

One-shot Learning with Python will guide you through the exploration and design of deep learning models that can obtain information about an object from one or just a few training samples. The course begins with an overview of deep learning and one-shot learning and then introduces you to the different methods you can use to achieve it, such as deep learning architectures and probabilistic models. Once you've got to grips with the core principles, you'll explore real-world examples and implementations of one-shot learning using PyTorch 1.x on datasets such as Omniglot and MiniImageNet. Finally, you'll explore generative modeling-based methods and discover the key considerations for building systems that exhibit human-level intelligence.

By the end of this course, you'll be well-versed with the different one- and few-shot learning methods and be able to use them to build your own deep learning models.

Working in a hands-on learning environment, led by Python expert instructor, students will learn about and explore:

* Learn how you can speed up the deep learning process with one-shot learning
* Use Python and PyTorch to build state-of-the-art one-shot learning models
* Explore architectures such as Siamese networks, memory-augmented neural networks, model-agnostic meta-learning, and discriminative k-shot learning.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get to grips with the fundamental concepts of one- and few-shot learning
* Work with different deep learning architectures for one-shot learning
* Understand when to use one-shot and transfer learning, respectively
* Study the Bayesian network approach for one-shot learning
* Implement one-shot learning approaches based on metrics, models, and optimization in PyTorch
* Discover different optimization algorithms that help to improve accuracy even with smaller volumes of data
* Explore various one-shot learning architectures based on classification and regression

**Audience & Pre-Requisites**

This course is for readers want to Get to grips with building powerful deep learning models using PyTorch and scikit-learn

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

1. **Section 1: One-shot Learning Introduction**

* Section 1: One-shot Learning Introduction

1. **Introduction to One-shot Learning**

* Introduction to One-shot Learning
* Technical requirements
* The human brain – overview
* Machine learning – historical overview
* One-shot learning – overview
* Setting up your environment
* Coding exercise

1. **Section 2: Deep Learning Architectures**

* Section 2: Deep Learning Architectures

1. **Metrics-Based Methods**

* Metrics-Based Methods
* Technical requirements
* Parametric methods – an overview
* Understanding Siamese networks
* Understanding matching networks
* Coding exercise

1. **5Model-Based Methods**

* Model-Based Methods
* Technical requirements
* Understanding Neural Turing Machines
* Memory-augmented neural networks
* Understanding meta networks
* Coding exercises

1. **Optimization-Based Methods**

* Optimization-Based Methods
* Technical requirements
* Overview of gradient descent
* Understanding model-agnostic meta-learning
* Understanding LSTM meta-learner

1. **Section 3: Other Methods and Conclusion**

* Section 3: Other Methods and Conclusion

1. **Generative Modeling-Based Methods**

* Generative Modeling-Based Methods
* Technical requirements
* Overview of Bayesian learning
* Understanding directed graphical models
* Overview of probabilistic methods
* Bayesian program learning
* Discriminative k-shot learning

1. **Conclusions and Other Approaches**

* Conclusions and Other Approaches
* Recent advancements
* Related fields
* Application

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with Go

**Course Snapshot**

* **Course:** Machine Learning with Go
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning with Go for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Get basic level of understanding when it comes to the Machine Learning (ML) development lifecycle, will introduce Go ML libraries and then will exemplify common ML methods such as Classification, Regression, and Clustering
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning is an essential part of today's data-driven world and is extensively used across industries, including financial forecasting, robotics, and web technology. This course will teach you how to efficiently develop machine learning applications in Go. The course starts with an introduction to machine learning and its development process, explaining the types of problems that it aims to solve and the solutions it offers. It then covers setting up a frictionless Go development environment, including running Go interactively with Jupyter notebooks. Finally, common data processing techniques are introduced. The course then teaches the reader about supervised and unsupervised learning techniques through worked examples that include the implementation of evaluation metrics. These worked examples make use of the prominent open-source libraries GoML and Gonum. The course also teaches readers how to load a pre-trained model and use it to make predictions. It then moves on to the operational side of running machine learning applications: deployment, Continuous Integration, and helpful advice for effective logging and monitoring. At the end of the course, readers will learn how to set up a machine learning project for success, formulating realistic success criteria and accurately translating business requirements into technical ones.

Working in a hands-on learning environment, led by Machine Learning with Go instructor, students will learn about and explore:

* Your handy guide to building machine learning workflows in Go for real-world scenarios
* Build predictive models using the popular supervised and unsupervised machine learning techniques
* Learn all about deployment strategies and take your ML application from prototype to production ready.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the types of problem that machine learning solves, and the various approaches
* Import, pre-process, and explore data with Go to make it ready for machine learning algorithms
* Visualize data with gonum/plot and Gophernotes
* Diagnose common machine learning problems, such as overfitting and underfitting
* Implement supervised and unsupervised learning algorithms using Go libraries
* Build a simple web service around a model and use it to make predictions

**Audience & Pre-Requisites**

This course is for readers want to Get basic level of understanding when it comes to the Machine Learning (ML) development lifecycle, will introduce Go ML libraries and then will exemplify common ML methods such as Classification, Regression, and Clustering

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

1. **Introducing Machine Learning with Go**

* Introducing Machine Learning with Go
* What is ML?
* Types of ML algorithms
* Why write ML applications in Go?
* ML development life cycle

1. **Setting Up the Development Environment**

* Setting Up the Development Environment
* Installing Go
* Running Go interactively with gophernotes
* Example – the most common phrases in positive and negative reviews
* Example – exploring body mass index data with gonum/plot
* Example – preprocessing data with Gota

1. **Supervised Learning**

* Supervised Learning
* Classification
* Regression

1. **Unsupervised Learning**

* Unsupervised Learning
* Clustering
* Principal component analysis

1. **Using Pretrained Models**

* Using Pretrained Models
* How to restore a saved GoML model
* Deciding when to adopt a polyglot approach
* Example – invoking a Python model using os/exec
* Example – invoking a Python model using HTTP
* Example – deep learning using the TensorFlow API for Go

1. **Deploying Machine Learning Applications**

* Deploying Machine Learning Applications
* The continuous delivery feedback loop
* Deployment models for ML applications

1. **Conclusion - Successful ML Projects**

* Conclusion - Successful ML Projects
* When to use ML
* Typical stages in a ML project
* When to combine ML with traditional code

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Applied SQL Data Analytics

**Course Snapshot**

* **Course:** Applied SQL Data Analytics
* **Duration:** 2 days
* **Skill-level**: Foundation-level Applied SQL Data Analytics skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Cut through the noise and get real results with a step-by-step approach to learning about SQL data analysis
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

You already know that you want to learn data analysis with SQL, and a smarter way to learn is to learn by doing. The Applied SQL Data Analytics Workshop focuses on building up your practical skills so that you can navigate and compose custom reports like an expert data analyst. You'll learn from real examples that lead to real results. Throughout The Applied SQL Data Analytics Workshop, you'll take an engaging step-by-step approach to understand data analytics with SQL. You won't have to sit through any unnecessary theory. You can jump into a single exercise each day if you're short on time, or you can spend an entire weekend tinkering with SQLAlchemy and Python. It's your choice. Learning on your terms, you'll build up and reinforce key skills in a way that feels rewarding. Every physical print copy of The Applied SQL Data Analytics Workshop unlocks access to the interactive edition. With videos detailing all exercises and activities, you'll always have a guided solution. You can also benchmark yourself against assessments, track progress, and receive content updates. You'll even earn a secure credential that you can share and verify online upon completion. It's a premium learning experience that's included with your printed copy. To redeem, follow the instructions located at the start of your course. Fast-paced and direct, The Applied SQL Data Analytics Workshop is the ideal companion for SQL beginners. You'll perform SQL queries like a professional data scientist, learning along the way. This process means that you'll find that your new skills stick, embedded as best practice. A solid foundation for the years ahead..

Working in a hands-on learning environment, led by Applied SQL Data Analytics instructor, students will learn about and explore:

* Ideal for SQL beginners who are getting started with data analytics for the first time
* A step-by-step SQL tutorial with exercises and activities that help build key skills
* Structured to let you progress at your own pace, on your own terms
* Use your physical print copy to redeem free access to the online interactive edition

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Experiment with data analytics using basic and advanced queries
* Learn data interpretation through descriptive statistics and aggregate functions
* Export data from outside sources using powerful SQL queries
* Discover how to work with and manipulate data using SQL joins and constraints
* Speed up your data analysis workflow by automating tasks and optimizing queries
* Discover different advanced analytics techniques, including geospatial and text analysis

**Audience & Pre-Requisites**

This course is for readers want to Cut through the noise and get real results with a step-by-step approach to learning about SQL data analysis

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

1. **Introduction to SQL for Analytics**

* Introduction
* The World of Data
* Methods of Descriptive Statistics
* Statistical Significance Testing
* Relational Databases and SQL
* Basic Data Types of SQL
* Reading Tables: The SELECT Query
* Creating Tables
* Updating Tables
* Deleting Data and Tables
* SQL and Analytics

1. **SQL for Data Preparation**

* Introduction
* Assembling Data
* Transforming Data

1. **Aggregate and Window Functions**

* Introduction
* Aggregate Functions
* Aggregate Functions with GROUP BY
* The HAVING Clause
* Using Aggregates to Clean Data and Examine Data Quality
* Window Function
* Statistics with Window Functions

1. **Importing and Exporting Data**

* Introduction
* The COPY Command
* Using R with Our Database
* Using Python with Our Database
* Best Practices for Importing and Exporting Data

1. **Analytics Using Complex Data Types**

* Introduction
* Date and Time Data Types for Analysis
* Performing Geospatial Analysis in PostgreSQL
* Using Array Data Types in PostgreSQL
* Using JSON Data Types in PostgreSQL
* Text Analytics Using PostgreSQL

1. **Performant SQL**

* Introduction
* Database Scanning Methods
* Performant Joins
* Functions and Triggers

1. **The Scientific Method and Applied Problem Solving**

* Introduction
* Case Study

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Supervised Learning

**Course Snapshot**

* **Course:** Supervised Learning
* **Duration:** 2 days
* **Skill-level**: Foundation-level Supervised Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those Cut through the noise and get real results with a step-by-step approach to understanding supervised learning algorithms
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

You already know you want to understand supervised learning, and a smarter way to do that is to learn by doing. The Supervised Learning Workshop focuses on building up your practical skills so that you can deploy and build solutions that leverage key supervised learning algorithms. You'll learn from real examples that lead to real results. Throughout The Supervised Learning Workshop, you'll take an engaging step-by-step approach to understand supervised learning. You won't have to sit through any unnecessary theory. If you're short on time you can jump into a single exercise each day or spend an entire weekend learning how to predict future values with auto regressors. It's your choice. Learning on your terms, you'll build up and reinforce key skills in a way that feels rewarding. Every physical print copy of The Supervised Learning Workshop unlocks access to the interactive edition. With videos detailing all exercises and activities, you'll always have a guided solution. You can also benchmark yourself against assessments, track progress, and receive content updates. You'll even earn a secure credential that you can share and verify online upon completion. It's a premium learning experience that's included with your printed copy. To redeem, follow the instructions located at the start of your course. Fast-paced and direct, The Supervised Learning Workshop is the ideal companion for those with some Python background who are getting started with machine learning. You'll learn how to apply key algorithms like a data scientist, learning along the way. This process means that you'll find that your new skills stick, embedded as best practice. A solid foundation for the years ahead.

Working in a hands-on learning environment, led by Supervised Learning instructor, students will learn about and explore:

* Ideal for those getting started with machine learning for the first time
* A step-by-step machine learning tutorial with exercises and activities that help build key skills
* Structured to let you progress at your own pace, on your own terms
* Use your physical print copy to redeem free access to the online interactive edition

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get to grips with the fundamental of supervised learning algorithms
* Discover how to use Python libraries for supervised learning
* Learn how to load a dataset in pandas for testing
* Use different types of plots to visually represent the data
* Distinguish between regression and classification problems
* Learn how to perform classification using K-NN and decision trees

**Audience & Pre-Requisites**

This course is for readers want to Cut through the noise and get real results with a step-by-step approach to understanding supervised learning algorithms

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

* **Fundamentals of Supervised** Learning Algorithms
* Fundamentals of Supervised Learning Algorithms
* Introduction
* Python Packages and Modules
* Data Quality Considerations

1. **Exploratory Data Analysis and Visualization**

* Introduction
* Exploratory Data Analysis (EDA)
* Summary Statistics and Central Values
* Missing Values
* Distribution of Values
* Relationships within the Data

1. **Linear Regression**

* Introduction
* Regression and Classification Problems
* Linear Regression
* Multiple Linear Regression

1. **Autoregression**

* Introduction
* Autoregression Models

1. **Classification Techniques**

* Introduction
* Ordinary Least Squares as a Classifier
* Logistic Regression
* Classification Using K-Nearest Neighbors
* Classification Using Decision Trees
* Artificial Neural Networks

1. **Ensemble Modeling**

* Introduction
* One-Hot Encoding
* Overfitting and Underfitting
* Bagging
* Bootstrapping
* Boosting
* Stacking

1. **Model Evaluation**

* Introduction
* Importing the Modules and Preparing Our Dataset
* Evaluation Metrics
* Splitting a Dataset
* Performance Improvement Tactics

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Supervised Machine Learning with Python

**Course Snapshot**

* **Course:** Supervised Machine Learning with Python
* **Duration:** 1 days
* **Skill-level**: Foundation-level Supervised Machine Learning with Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those want machine to think for itself!
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Supervised machine learning is used in a wide range of sectors (such as finance, online advertising, and analytics) because it allows you to train your system to make pricing predictions, campaign adjustments, customer recommendations, and much more while the system self-adjusts and makes decisions on its own. As a result, it's crucial to know how a machine “learns” under the hood. This course will guide you through the implementation and nuances of many popular supervised machine learning algorithms while facilitating a deep understanding along the way. You’ll embark on this journey with a quick overview and see how supervised machine learning differs from unsupervised learning. Next, we explore parametric models such as linear and logistic regression, non-parametric methods such as decision trees, and various clustering techniques to facilitate decision-making and predictions. As we proceed, you'll work hands-on with recommender systems, which are widely used by online companies to increase user interaction and enrich shopping potential. Finally, you’ll wrap up with a brief foray into neural networks and transfer learning. By the end of this course, you’ll be equipped with hands-on techniques and will have gained the practical know-how you need to quickly and powerfully apply algorithms to new problems.

Working in a hands-on learning environment, led by Supervised Machine Learning with Python instructor, students will learn about and explore:

* Delve into supervised learning and grasp how a machine learns from data
* Implement popular machine learning algorithms from scratch, developing a deep understanding along the way
* Explore some of the most popular scientific and mathematical libraries in the Python language

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Crack how a machine learns a concept and generalize its understanding to new data
* Uncover the fundamental differences between parametric and non-parametric models
* Implement and grok several well-known supervised learning algorithms from scratch
* Work with models in domains such as ecommerce and marketing
* Expand your expertise and use various algorithms such as regression, decision trees, and clustering
* Build your own models capable of making predictions
* Delve into the most popular approaches in deep learning such as transfer learning and neural networks

**Audience & Pre-Requisites**

This course is for readers want to machine to think for itself!

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* For readers with existing programming skills.

**Course Agenda / Topics**

1. **First Step Towards Supervised Learning**

* First Step Towards Supervised Learning
* Technical requirements
* An example of supervised learning in action
* Setting up the environment
* Supervised learning
* Hill climbing and loss functions
* Hill climbing and descent
* Model evaluation and data splitting

1. **Implementing Parametric Models**

* Implementing Parametric Models
* Technical requirements
* Parametric models
* Implementing linear regression from scratch
* Logistic regression models
* Implementing logistic regression from scratch
* The pros and cons of parametric models

1. **Working with Non-Parametric Models**

* Working with Non-Parametric Models
* Technical requirements
* The bias/variance trade-off
* Introduction to non-parametric models and decision trees
* Decision trees
* Implementing a decision tree from scratch
* Various clustering methods
* Implementing KNNs from scratch
* Non-parametric models – pros/cons

1. **Advanced Topics in Supervised Machine Learning**

* Advanced Topics in Supervised Machine Learning
* Technical requirements
* Recommended systems and an introduction to collaborative filtering
* Matrix factorization
* Matrix factorization in Python
* Content-based filtering
* Neural networks and deep learning
* Neural networks
* Using transfer learning

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with C++.

**Course Snapshot**

* **Course:** Machine Learning with C++
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning with C++ skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Implement supervised and unsupervised machine learning algorithms using C++ libraries such as PyTorch C++ API, Caffe2, Shogun, Shark-ML, mlpack, and dlib with the help of real-world examples and datasets
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

C++ can make your machine learning models run faster and more efficiently. This handy guide will help you learn the fundamentals of machine learning (ML), showing you how to use C++ libraries to get the most out of your data. This coursemakes machine learning with C++ for beginners easy with its example-based approach, demonstrating how to implement supervised and unsupervised ML algorithms through real-world examples.This course will get you with tuning and optimizing a model for different use cases, assisting you with model selection and the measurement of performance. You’ll cover techniques such as product recommendations, ensemble learning, and anomaly detection using modern C++ libraries such as PyTorch C++ API, Caffe2, Shogun, Shark-ML, mlpack, and dlib. Next, you’ll explore neural networks and deep learning using examples such as image classification and sentiment analysis, which will help you solve various problems. Later, you’ll learn how to handle production and deployment challenges on mobile and cloud platforms, before discovering how to export and import models using the ONNX format. By the end of this C++ course, you will have real-world machine learning and C++ knowledge, as well as the skills to use C++ to build powerful ML systems.

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Become familiar with data processing, performance measuring, and model selection using various C++ libraries
* Implement practical machine learning and deep learning techniques to build smart models
* Deploy machine learning models to work on mobile and embedded devices

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Explore how to load and preprocess various data types to suitable C++ data structures
* Employ key machine learning algorithms with various C++ libraries
* Understand the grid-search approach to find the best parameters for a machine learning model
* Implement an algorithm for filtering anomalies in user data using Gaussian distribution
* Improve collaborative filtering to deal with dynamic user preferences
* Use C++ libraries and APIs to manage model structures and parameters
* Implement a C++ program to solve image classification tasks with LeNet architecture

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Implement supervised and unsupervised machine learning algorithms using C++ libraries such as PyTorch C++ API, Caffe2, Shogun, Shark-ML, mlpack, and dlib with the help of real-world examples and datasets

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Section 1: Overview of** **Machine Learning**
2. **Introduction to Machine Learning with C++**

* Introduction to Machine Learning with C++
* Understanding the fundamentals of ML
* An overview of linear algebra
* An overview of linear regression

1. **Data Processing**

* Data Processing
* Technical requirements
* Parsing data formats to C++ data structures
* Initializing matrix and tensor objects from C++ data structures
* Manipulating images with the OpenCV and Dlib libraries
* Transforming images into matrix or tensor objects of various libraries
* Normalizing data

1. **Measuring Performance and Selecting Models**

* Technical requirements
* Performance metrics for ML models
* Understanding the bias and variance characteristics
* Model selection with the grid search technique

1. **Section 2: Machine Learning Algorithms**
2. **Clustering**

* Technical requirements
* Measuring distance in clustering
* Types of clustering algorithms
* Examples of using the Shogun library for dealing with the clustering task samples
* Examples of using the Shark-ML library for dealing with the clustering task samples
* Examples of using the Dlib library for dealing with the clustering task samples
* Plotting data with C++

1. **Anomaly Detection**

* Technical requirements
* Exploring the applications of anomaly detection
* Learning approaches for anomaly detection
* Examples of using different C++ libraries for anomaly detection

1. **Dimensionality Reduction**

* Technical requirements
* An overview of dimension reduction methods
* Exploring linear methods for dimension reduction
* Exploring non-linear methods for dimension reduction
* Understanding dimension reduction algorithms with various С++ libraries

1. **Classification**

* Technical requirements
* An overview of classification methods
* Exploring various classification methods
* Examples of using C++ libraries for dealing with the classification task

1. **Recommender Systems**

* Technical requirements
* An overview of recommender system algorithms
* Understanding collaborative filtering method details
* Examples of item-based collaborative filtering with C++

1. **Ensemble Learning**

* Ensemble Learning
* Technical requirements
* An overview of ensemble learning
* Examples of using C++ libraries for creating ensembles

1. **Section 3: Advanced Examples**
2. **Neural Networks for Image Classification**

* Technical requirements
* An overview of neural networks
* Delving into convolutional networks
* What is deep learning?
* Examples of using C++ libraries to create neural networks
* Understanding image classification using the LeNet architecture

1. **Sentiment Analysis with Recurrent Neural Networks**

* Sentiment Analysis with Recurrent Neural Networks
* Technical requirements
* An overview of the RNN concept
* Training RNNs using the concept of backpropagation through time
* Exploring RNN architectures
* Understanding natural language processing with RNNs
* Sentiment analysis example with an RNN

1. **Section 4: Production and Deployment Challenges**
2. **Exporting and Importing Models**

* Technical requirements
* ML model serialization APIs in C++ libraries
* Delving into ONNX format

1. **Deploying Models on Mobile and Cloud Platforms**

* Technical requirements
* Image classification on Android mobile
* Machine learning in the cloud – using Google Compute Engine

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Natural Language Processing for Hackers

Course Snapshot

* **Course:** Natural Language Processing for Hackers
* **Duration:** 6 days
* **Skill-level**: Foundation-level Natural-Language-Processing skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to know the techniques to analyze, interpret, and create human-understandable text and speech. Advances in machine learning have pushed NLP to new levels of accuracy and uncanny realism
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Natural Language Processing for Hackers** covers NLP end-to-end, giving you the skills and techniques that allow your computers to speak human. Unlike many research-oriented courses that use the kind of clean datasets you would never find in the real world; this practical guide takes on NLP as you’ll actually use it. You’ll learn the key concepts of NLP by coding your own tools and projects, from a text analysis service right up to a full-featured chatbot. Everything is written in concise, easy-to-read Python code to ensure you’ll grok the most important aspects of Natural Language Processing. When you’re done, you will be able to apply the complete range of NLP techniques to build practical applications—even with messy real-world data.

Working in a hands-on learning environment, led by our Natural Language Processing expert instructor, students will learn about and explore:

* You’ll learn the key concepts of NLP by coding your own tools and projects, from a text analysis service right up to a full-featured chatbot. Everything is written in concise, easy-to-read Python code to ensure
* you’ll grok the most important aspects of Natural Language Processing.
* you will be able to apply the complete range of NLP techniques to build practical applications—even with messy real-world data.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Constructing your own Text Analysis engine
* Building a Twitter listener that performs Sentiment Analysis on a certain subject
* Assembling your own NLP toolbox, complete with Part of Speech Tagger, Shallow Parser, Named Entity Extractor, and Dependency Parsers
* Cleaning and standardizing messy datasets

Audience & Pre-Requisites

This course is geared for attendees with Python skills who wish to who wants to know the techniques to analyze, interpret, and create human-understandable text and speech

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Attendees requires familiarity with Python, but no prior knowledge of natural language processing or machine learning is necessary.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

**PART: INTRODUCTION TO NLTK**

1. **NLTK FUNDAMENTALS**

* Installing NLTK
* Splitting Text
* Building a vocabulary
* Fun with Bigrams and Trigrams
* Part of Speech Tagging
* Named Entity Recognition

1. **GETTING STARTED WITH WORDNET**

* Wordnet Structure
* Lemma Operations

1. **LEMMATIZING AND STEMMING**

* How stemmers work
* How lemmatizes work

**PART: CREATE A TEXT ANALYSIS SERVICE**

1. **INTRODUCTION TO MACHINE LEARNING**

* A Practical Machine Learning Example

1. **GETTING STARTED WITH SCIKIT-LEARN**

* Installing Scikit-Learn and building a dataset
* Training a Scikit-Learn Model
* Making Predictions

1. **FINDING THE DATA**

* Existing corpora
* Ideas for Gathering Data
* Getting the Data

1. **LEARNING TO CLASSIFY TEXT**

* Text Feature Extractor
* Scikit-Learn Feature Extraction
* Text Classification with Naive Bayes

1. **PERSISTING MODELS**
2. **BUILDING THE API**

* Building a Flask API Deploy to Heroku

**PART: CREATE A SOCIAL MEDIA MONITORING SERVICE**

1. **BASICS OF SENTIMENT ANALYSIS**

* Be Aware of Negations
* Machine Learning doesn't get Humor
* Multiple and Mixed Sentiments
* Non-Verbal Communication

1. **TWITTER SENTIMENT DATA**

* Twitter Corpora
* Other Sentiment Analysis Corpora
* Building a Tweets Dataset
* Sentiment Analysis - A First Attempt
* Better Tokenization

1. **FINE TUNING**

* Try a different classifier
* Use Ngrams Instead of Words
* Using a Pipeline
* Cross Validation
* Grid Search
* Picking the Best Tokenizer

1. **BUILDING THE TWITTER LISTENER**
2. **CLASSIFICATION METRICS**

* Binary Classification

1. **MULTI-CLASS METRICS**

* The Confusion Matrix

1. **BUILD YOUR OWN PART-0F-SPEECH TAGGER**

* Part-Of-Speech Corpora
* Building Toy Models
* About Feature Extraction
* Using the NLTK Base Classes
* Writing the Feature Extractor
* Training the Tagger
* Out-Of-Core Learning

1. **BUILD A CHUNKER**

* IOB Tagging
* Implementing the Chunk Parser
* Chunker Feature Detection

1. **BUILD A NAMED ENTITY EXTRACTOR**

* NER Corpora
* The Groningen Meaning Bank Corpus
* Feature Detection
* NER Training

**PART: BUILD YOUR OWN CHATBOT ENGINE**

1. **GENERAL ARCHITECTURE**

* Train the Platform via Examples
* Action Handlers

1. **BUILDING THE CORE**

* Chatbot Base Class and Training Set
* Training the Chatbot Everything together

1. **MOVIEBOT**

* The Movie DB API
* Small-Talk Handlers
* Simple Handlers
* Execution Handlers

1. **MOVIEBOT ON FACEBOOK**

* Installing ngrok Setting up Facebook

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with ML.NET.

**Course Snapshot**

* **Course:** Machine Learning with ML.NET
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning with ML.NETskills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who are intending to Create, train, and evaluate various machine learning models such as regression, classification, and clustering using ML.NET, Entity Framework, and ASP.NET Core
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine learning (ML) is widely used in many industries such as science, healthcare, and research and its popularity is only growing. In March 2018, Microsoft introduced ML.NET to help .NET enthusiasts in working with ML. With this course, you’ll explore how to build ML.NET applications with the various ML models available using C# code. The course starts by giving you an overview of ML and the types of ML algorithms used, along with covering what ML.NET is and why you need it to build ML apps. You’ll then explore the ML.NET framework, its components, and APIs. The course will serve as a practical guide to helping you build smart apps using the ML.NET library. You’ll gradually become well versed in how to implement ML algorithms such as regression, classification, and clustering with real-world examples and datasets. Each lesson will cover the practical implementation, showing you how to implement ML within .NET applications. You’ll also learn to integrate TensorFlow in ML.NET applications. Later you’ll discover how to store the regression model housing price prediction result to the database and display the real-time predicted results from the database on your web application using ASP.NET Core Blazor and SignalR. By the end of this course, you’ll have learned how to confidently perform basic to advanced-level machine learning tasks in ML.NET.

Working in a hands-on learning environment, led by our Machine Learning with ML.NET expert instructor, students will learn about and explore:

* Get well-versed with the ML.NET framework and its components and APIs using practical examples
* Learn how to build, train, and evaluate popular machine learning algorithms with ML.NET offerings
* Extend your existing machine learning models by integrating with TensorFlow and other libraries

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the framework, components, and APIs of ML.NET using C#
* Develop regression models using ML.NET for employee attrition and file classification
* Evaluate classification models for sentiment prediction of restaurant reviews
* Work with clustering models for file type classifications
* Use anomaly detection to find anomalies in both network traffic and login history
* Work with ASP.NET Core Blazor to create an ML.NET enabled web application
* Integrate pre-trained TensorFlow and ONNX models in a WPF ML.NET application for image classification and object detection

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Create, train, and evaluate various machine learning models such as regression, classification, and clustering using ML.NET, Entity Framework, and ASP.NET Core.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Section 1: Fundamentals of Machine Learning and ML.NET**

* Section 1: Fundamentals of Machine Learning and ML.NET

1. **Getting Started with Machine Learning and ML.NET**

* Getting Started with Machine Learning and ML.NET
* The importance of learning about machine learning today
* The model building process
* Exploring types of learning
* Exploring various machine learning algorithms
* What is ML.NET?

1. **Setting Up the ML.NET Environment**

* Setting Up the ML.NET Environment
* Setting up your development environment
* Creating your first ML.NET application
* Evaluating the model

1. **Section 2: ML.NET Models**

* Section 2: ML.NET Models

1. **Regression Model**

* Regression Model
* Breaking down regression models
* Creating the linear regression application
* Creating the logistic regression application
* Evaluating a regression model

1. **Classification Model**

* Classification Model
* Breaking down classification models
* Creating a binary classification application
* Creating a multi-class classification application
* Evaluating a classification model

1. **Clustering Model**

* Clustering Model
* Breaking down the k-means algorithm
* Creating the clustering application
* Evaluating a k-means model

1. **Anomaly Detection Model**

* Anomaly Detection Model
* Breaking down anomaly detection
* Creating a time series application
* Creating an anomaly detection application
* Evaluating a randomized PCA model

1. **Matrix Factorization Model**

* Matrix Factorization Model
* Breaking down matrix factorizations
* Creating a matrix factorization application
* Evaluating a matrix factorization model

1. **Section 3: Real-World** **Integrations with ML.NET**

* Section 3: Real-World Integrations with ML.NET

1. **Using ML.NET with .NET Core and Forecasting**

* Using ML.NET with .NET Core and Forecasting
* Breaking down the .NET Core application architecture
* Creating the stock price estimator application
* Exploring additional production application enhancements

1. **Using ML.NET with ASP.NET Core**

* Using ML.NET with ASP.NET Core
* Breaking down ASP.NET Core
* Creating the file classification web application
* Exploring additional ideas for improvements

1. **Using ML.NET with UWP**

* Using ML.NET with UWP
* Breaking down the UWP architecture
* Creating the web browser classification application
* Additional ideas for improvements

1. **Section 4: Extending ML.NET**

* Section 4: Extending ML.NET

1. **Training and Building Production Models**

* Training and Building Production Models
* Investigating feature engineering
* Obtaining training and testing datasets
* Creating your model-building pipeline

1. **Using TensorFlow with ML.NET**
2. **Using ONNX with ML.NET**

* Using ONNX with ML.NET
* Breaking down ONNX and YOLO
* Creating the ONNX object detection application
* Exploring additional production application enhancements

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with TensorFlow

**Course Snapshot**

* **Course:** Machine Learning with TensorFlow
* **Duration:** 5 days
* **Skill-level**: Foundation-level Machine Learning with TensorFlow skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to learn the foundational concepts of machine learning, and how to utilize the TensorFlow library to rapidly build powerful ML models
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

This fully revised edition of **Machine Learning with TensorFlow** teaches you the foundational concepts of machine learning, and how to utilize the TensorFlow library to rapidly build powerful ML models. You’ll learn the basics of regression, classification, and clustering algorithms, applying them to solve real-world challenges such as call center volume prediction and sentiment analysis of movie reviews. Once you’ve mastered core ML concepts, you’ll move on to the money lessons: exploring cutting-edge neural network techniques such as deep speech classifiers, facial identification, and auto-encoding with CIFAR-10. Digest this course, and you’ll be able to start modelling your everyday problems as automated machine learning tasks.

Working in a hands-on learning environment, led by our Machine Learning with TensorFlow expert instructor, students will learn about and explore:

* how to utilize the TensorFlow library to rapidly build powerful ML models
* You’ll learn the basics of regression, classification, and clustering algorithms, applying them to solve real-world challenges such as call center volume prediction and sentiment analysis of movie reviews. Once you’ve mastered core ML concepts
* you’ll be able to start modelling your everyday problems as automated machine learning tasks.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Matching your tasks to the right machine-learning or deep-learning approach
* Visualizing algorithms with Tensor Board
* Sharing your results with collaborators using other frameworks
* Understanding and using neural networks
* Reproducing and employing predictive science
* Downloadable Jupyter Notebooks for all examples
* Questions to test your knowledge
* Examples use the super-stable 1.14.1 branch of TensorFlow

**Audience & Pre-Requisites**

This course is geared for attendees who wants to learn the foundational concepts of machine learning, and how to utilize the TensorFlow library to rapidly build powerful ML models

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, and Machine Learning knowledge
* Good foundational mathematics or logic skills
* experienced with Python and algebraic concepts like vectors and matrices.

**Course Agenda / Topics**

**PART 1: YOUR MACHINE-LEARNING RIG**

1. **A machine-learning odyssey**

* Machine-learning fundamentals
* Data representation and features
* Distance metrics
* Types of learning
* TensorFlow

1. **TensorFlow essentials**

* Ensuring that TensorFlow works
* Representing tensors
* Creating operators
* Executing operators with sessions
* Understanding code as a graph
* Writing code in Jupyter
* Using variables
* Saving and loading variables
* Visualizing data using TensorBoard
* Putting it all together: The TensorFlow System Architecture and. API

**PART 2: CORE LEARNING ALGORITHMS**

1. **Linear regression and beyond**

* Formal notation
* Linear regression
* Polynomial model
* Regularization
* Application of linear regression

1. **Using regression for call center volume prediction**

* What is 3-1-1?
* Cleaning the data for regression
* What’s in a bell curve: predicting Gaussian distributions
* Training your call prediction regressor
* Visualizing the results and plotting the error
* Regularization and train test splits

1. **A gentle introduction to classification**

* Formal notation
* Measuring performance
* Using linear regression for classification
* Using logistic regression
* Multiclass classifier
* Application of classification

1. **Sentiment classification: Netflix large movie-review dataset**

* The Bag of Words model
* Analysis on your Bag of Words
* Building a sentiment classifier using logistic regression
* Making predictions using your sentiment classifier
* Measuring the effectiveness of your classifier
* Creating the softmax-regression sentiment classifier
* Submit your results to Kaggle

1. **Automatically clustering data**

* Traversing files in TensorFlow
* Extracting features from audio
* K-means clustering
* Audio segmentation
* Clustering using a self-organizing map
* Application of clustering

1. **Inferring user activity from Android accelerometer data**

* The user activity from walking dataset
* Clustering similar participants based on jerk magnitudes
* Different classes of user activity for a single participant: climbing, standing, walking, talking, and working

1. **Hidden Markov models**

* Example of a not-so-interpretable model
* Markov model
* Hidden Markov model
* Forward algorithm
* Viterbi decoding
* Uses of hidden Markov models
* Application of hidden Markov models

1. **Part of speech tagging and word sense disambiguation**

* Review the HMM example: rainy or sunny and what it’s actually doing
* Part-of-speech tagging
* Algorithms for building the Hidden Markov Model for PoS disamguiation
* Running the HMM and evaluating its output
* Getting more training data using the Brown corpus
* Defining error bars and metrics for PoS tagging

**PART 3: THE NEURAL NETWORK PARADIGM**

1. **A peek into autoencoders**

* Neural networks
* Autoencoders
* Batch training
* Working with images
* Application of autoencoders

1. **Applying autoencoders: the CIFAR- image dataset**
2. **Reinforcement learning**
3. **Convolutional neural networks**
4. **Building a real-world CNN: VGG-Face and VGG-Face Lite**
5. **Recurrent neural networks**
6. **LSTMs and automatic speech recognition**
7. **Sequence-to-sequence models for chatbots**
8. **Utility landscape**

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Zero to AI

**Course Snapshot**

* **Course:** Zero to AI
* **Duration:** 3 days
* **Skill-level**: Foundation-level Zero to AI for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for business leaders, entrepreneurs, and decision makers as well as technology implementers looking for a big-picture view of AI.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Zero to AI** teaches business leaders, entrepreneurs, and decision makers how to improve the success and efficiency of their businesses by taking advantage of state-of-the-art AI technologies. After a brief introduction to artificial intelligence, you’ll explore examples that demonstrate how you can use AI for analyzing business data, predicting customer buying trends, deciphering text and images, and much more. The course is filled with extensive, real-world case studies. As you go, you’ll learn how Google applied AI models to improve on century-old engineering rules to save energy (and money) in its data centers. You’ll look under the hood of the models that power Netflix’s video recommendations and see how they compare with the algorithms that Target uses to prepare their customized promotions. For each case study, the authors discuss the best plan of attack, the necessary resources, the possible risk factors, and likely business benefits of the specific AI application. When you’re done, you’ll have a complete roadmap for realizing the vast potential of AI in your own organization!

Working in a hands-on learning environment, led by our AI expert instructor, students will learn about and explore:

* Unprecedented access to raw data and affordable computing power, along with incredible advances in AI, put these smart, powerful, machine learning systems within reach of nearly any organization.
* Explore how businesses in every industry are using AI to streamline processes, personalize marketing, improve customer engagement, and grow their bottom lines.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

identifying opportunities for applying AI in your organization

Designing an AI strategy

Hiring AI talent

Managing an AI project

Using AI for boosting conversion rates

Content curation and community building with AI

Image classification and object recognition

Interesting real-world case studies from companies like Google, Square, Netflix, and Target

**Audience & Pre-Requisites**

This course is geared for for business leaders, entrepreneurs, and decision makers as well as technology implementers looking for a big-picture view of AI.

**Pre-Requisites:** Students should have

* No prior programming or machine learning knowledge is required.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **An introduction to Artificial Intelligence**

* The path to modern AI
* The engine of the AI revolution: Machine Learning
* What is Artificial Intelligence after all?
* Our teaching method

` Part 1: Understanding AI

1. **Artificial Intelligence for core business data**

* Unleashing AI on core business data
* Using AI with core business data
* Case studies
* Evaluating performance and risk

1. **AI for sales and marketing**

* Why AI for Sales and Marketing
* Predicting churning customers
* Using AI to boost conversion rates and upselling
* Automated customer segmentation
* Measuring performance
* Tying Machine Learning metrics to business outcomes and risks
* AI for sales and marketing case studies

1. **AI for media**

* Improving products with Computer Vision
* AI for image classification: What is Deep Learning?
* Small datasets and Transfer Learning
* Face recognition: teaching computers to recognize people
* Content generation and Style Transfer
* What to watch out for
* AI for audio
* Case study: optimizing agriculture with Deep Learning

1. **AI for natural language**

* The allure of Natural Language Understanding
* Breaking down Natural Language Processing: Measuring complexity
* Adding NLP capabilities to your organization
* overcome technology limitations
* Case study: Translated.com

1. **AI for content curation and community building**

* The curse of choice
* Driving engagement with Recommender Systems
* The wisdom of crowds: collaborative filtering
* Recommendations gone wrong
* Case study: How Netflix saves $1bn a year using Recommender Systems

1. **Ready - finding AI opportunities**

* Don’t fall for the hype - business-driven AI innovation
* Invention: scouting for AI opportunities
* Prioritization: Evaluating AI projects
* Validate: Analyzing risks
* Deconstructing an AI product
* Translating an AI project into ML-friendly terms

1. **Set - preparing data, technology and people**

* Data Strategy
* Where do I get data?
* How much data do I need?
* Data quality
* Recruiting an AI team

1. **Go - AI implementation strategy**

* Buying or building AI
* The Lean Strategy
* The virtuous cycle of AI
* Managing AI projects
* When AI fails

1. **What lies ahead**

* How AI threatens society
* 10.2 Opportunities for AI in society
* 10.3 Opportunities for AI in industries
* What about general AI?
* Closing thoughts

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Deep Learning with R

**Course Snapshot**

* **Course:** Deep Learning with R
* **Duration:** 2 days
* **Skill-level**: Foundation-level Deep Learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to use the powerful Keras library and its R language interface
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Deep Learning with R introduces the world of deep learning using the powerful Keras library and its R language interface. Initially written for Python as Deep Learning with Python by Keras creator and Google AI researcher François Chollet and adapted for R by RStudio founder J. J. Allaire, this course builds your understanding of deep learning through intuitive explanations and practical examples. You'll practice your new skills with R-based applications in computer vision, natural-language processing, and generative models.

Working in a hands-on learning environment, led by our Deep Learning with R expert instructor, students will learn about and explore:

* use the powerful Keras library and its R language interface.
* builds your understanding of deep learning through intuitive explanations and practical examples.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Deep learning from first principles
* Setting up your own deep-learning environment
* Image classification and generation
* Deep learning for text and sequences

**Audience & Pre-Requisites**

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* You'll need intermediate R programming skills.
* No previous experience with machine learning or deep learning is assumed.

**Course Agenda / Topics**

1. **What is deep learning?free**

* Artificial intelligence, machine learning, and deep learning
* Before deep learning: a brief history of machine learning
* Why deep learning? Why now?

1. **Before we begin: the mathematical building blocks of neural networksfree**

* A first look at a neural network
* Data representations for neural networks
* The gears of neural networks: tensor operations
* The engine of neural networks: gradient-based optimization
* Looking back at our first example

1. **Getting started with neural networksfree**

* Anatomy of a neural network
* Introduction to Keras
* Setting up a deep-learning workstation
* Classifying movie reviews: a binary classification example
* Classifying newswires: a multiclass classification example
* Predicting house prices: a regression example

1. **Fundamentals of machine learning**

* Four branches of machine learning
* Evaluating machine-learning models
* Data preprocessing, feature engineering, and feature learning
* Overfitting and underfitting
* The universal workflow of machine learning

1. **Deep learning for computer vision**

* Introduction to convnets
* Training a convnet from scratch on a small dataset
* Using a pretrained convnet
* Visualizing what convnets learn

1. **Deep learning for text and sequences**

* Working with text data
* Understanding recurrent neural networks
* Advanced use of recurrent neural networks
* Sequence processing with convnets

1. **Advanced deep-learning best practices**

* Going beyond the sequential model: the Keras functional API
* Inspecting and monitoring deep-learning models using Keras callba- acks and TensorBoard
* Getting the most out of your models

1. **Generative deep learning**

* Text generation with LSTM
* DeepDream
* Neural style transfer
* Generating images with variational autoencoders
* Introduction to generative adversarial networks

1. **Conclusions**

* Key concepts in review
* The limitations of deep learning
* The future of deep learning
* Staying up to date in a fast-moving field
* Final words

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

**Deep Learning with Python**

**Course Snapshot**

* **Course:** Deep Learning with Python
* **Duration:** 3 days
* **Skill-level**: Foundation-level Deep Learning with Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who want to build understand through practical examples and intuitive explanations that make the complexities of deep learning accessible and understandable.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Deep Learning** with Python introduces the field of deep learning using the Python language and the powerful Keras library. You’ll learn directly from the creator of Keras, François Cholet, building your understanding through intuitive explanations and practical examples. Updated from the original bestseller with over 50% new content, this edition includes new lessons, cutting-edge innovations, and coverage of the very latest deep learning tools. You'll explore challenging concepts and practice with applications in computer vision, natural-language processing, and generative models. By the time you finish, you'll have the knowledge and hands-on skills to apply deep learning in your own projects.

Working in a hands-on learning environment, led by our Deep Learning with Python expert instructor, students will learn about and explore:

* You'll explore challenging concepts and practice with applications in computer vision, natural-language processing, and generative models.
* By the time you finish, you'll have the knowledge and hands-on skills to apply deep learning in your own projects.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Deep learning from first principles
* Image-classification, imagine segmentation, and object detection
* Deep learning for natural language processing
* Timeseries forecasting
* Neural style transfer, text generation, and image generation

**Audience & Pre-Requisites**

This course is geared for attendees with Deep Learning with Python who wish to explore challenging concepts and practice with applications in computer vision, natural-language processing, and generative models. By the time you finish, you'll have the knowledge and hands-on skills to apply deep learning in your own projects.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Readers need intermediate Python skills.
* No previous experience with Keras, TensorFlow, or machine learning is required.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **What is deep learning?**

* Artificial intelligence, machine learning, and deep learning
* Before deep learning: a brief history of machine learning
* Why deep learning? Why now?

1. **The mathematical building blocks of neural networks**

* A first look at a neural network
* Data representations for neural networks
* The gears of neural networks: tensor operations
* The engine of neural networks: gradient-based optimization
* Looking back at our first example

1. **Introduction to Keras and TensorFlow**

* What’s TensorFlow?
* What’s Keras?
* Keras and TensorFlow: a brief history
* Setting up a deep-learning workspace
* First steps with TensorFlow
* Anatomy of a neural network: understanding core Keras APIs

1. **Getting started with neural networks: classification and regression**

* Classifying movie reviews: a binary classification example
* Classifying newswires: a multiclass classification example
* Predicting house prices: a regression example

1. **Fundamentals of machine learning**

* Generalization: the goal of machine learning
* Evaluating machine-learning models
* Improving model fit
* Improving generalization

1. **The universal workflow of machine learning**

* Define the task
* success
* Develop a model
* Deploy your model

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning.

**Course Snapshot**

* **Course:** Machine Learning
* **Duration:** 4 days
* **Skill-level**: Foundation-level machine learning skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to learn flexible Python programming language to build programs that implement algorithms for data classification, forecasting, recommendations, and higher-level features like summarization and simplification.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

A machine is said to learn when its performance improves with experience. Learning requires algorithms and programs that capture data and ferret out the interesting or useful patterns. Once the specialized domain of analysts and mathematicians, machine learning is becoming a skill needed by many. **Machine Learning** is a clearly written tutorial for developers. It avoids academic language and takes you straight to the techniques you'll use in your day-to-day work. Many (Python) examples present the core algorithms of statistical data processing, data analysis, and data visualization in code you can reuse. You'll understand the concepts and how they fit in with tactical tasks like classification, forecasting, recommendations, and higher-level features like summarization and simplification.

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* you'll use in your day-to-day work. Many (Python) examples present the core algorithms of statistical data processing, data analysis, and data visualization in code you can reuse.
* You'll understand the concepts and how they fit in with tactical tasks like classification, forecasting, recommendations, and higher-level features like summarization and simplification.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* A no-nonsense introduction
* Examples showing common ML tasks
* Everyday data analysis
* Implementing classic algorithms like Apriori and Adobos

**Audience & Pre-Requisites**

This course is geared introduces the day-to-day practice of machine learning, preparing you to successfully build and deploy powerful ML systems.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Readers need no prior experience with machine learning or statistical processing. Familiarity with Python is helpful.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Machine learning basics free**

* What is machine learning?
* Key terminology
* Key tasks of machine learning
* How to choose the right algorithm
* Steps in developing a machine learning application
* Why Python?
* Getting started with the NumPy library

1. **Classifying with k-Nearest Neighbors**

* Classifying with distance measurements
* Example: improving matches from a dating site with kNN
* Example: a handwriting recognition system

1. **Splitting datasets one feature at a time: decision trees**

* Tree construction
* Plotting trees in Python with Matplotlib annotations
* annotations
* Testing and storing the classifier
* Example: using decision trees to predict contact lens type

1. **Classifying with probability theory: naïve Bayes**

* Classifying with Bayesian decision theory
* Conditional probability
* Classifying with conditional probabilities
* Document classification with naïve Bayes
* Classifying text with Python
* Example: classifying spam email with naïve Bayes
* Example: using naïve Bayes to reveal local attitudes from personal ads

1. **Logistic regression**

* Classification with logistic regression and the sigmoid function: -: a tractable step function
* Using optimization to find the best regression coefficients
* Example: estimating horse fatalities from colic

1. **Support vector machines**

* Separating data with the maximum margin
* Finding the maximum margin
* Efficient optimization with the SMO algorithm
* Speeding up optimization with the full Platt SMO
* Using kernels for more complex data
* Example: revisiting handwriting classification

1. **Improving classification with the AdaBoost meta-algorithm**

* Classifiers using multiple samples of the dataset
* Train: improving the classifier by focusing on errors
* Creating a weak learner with a decision stump
* Implementing the full AdaBoost algorithm
* Test: classifying with AdaBoost
* Example: AdaBoost on a difficult dataset
* Classification imbalance

1. **Predicting numeric values: regression**

* Finding best-fit lines with linear regression
* Locally weighted linear regression
* Example: predicting the age of an abalone
* Shrinking coefficients to understand our data
* The bias/variance tradeoff
* Example: forecasting the price of LEGO sets

1. **Tree-based regression**

* Locally modeling complex data
* Building trees with continuous and discrete features
* Using CART for regression
* Tree pruning
* Model trees
* Example: comparing tree methods to standard regression
* Using Tkinter to create a GUI in Python

1. **Grouping unlabeled items using k-means clustering**

* The k-means clustering algorithm
* Improving cluster performance with postprocessing
* Bisecting k-means
* Example: clustering points on a map

1. **Association analysis with the Apriori algorithm**

* Association analysis
* The Apriori principle
* Finding frequent itemset with the Apriori algorithm
* Mining association rules from frequent item sets
* Example: uncovering patterns in congressional voting
* Example: finding similar features in poisonous mushrooms

1. **Efficiently finding frequent item sets with FP-growth**

* FP-trees: an efficient way to encode a dataset
* Build an FP-tree
* Mining frequent items from an FP-tree
* Example: finding co-occurring words in a Twitter feed
* Example: mining a clickstream from a news site

1. **Using principal component analysis to simplify data**

* Dimensionality reduction techniques
* Principal component analysis
* Example: using PCA to reduce the dimensionality of semiconductor - manufacturing data

1. **Simplifying data with the singular value decomposition**

* Applications of the SVD
* Matrix factorization
* SVD in Python
* Collaborative filtering–based recommendation engines
* Example: a restaurant dish recommendation engine
* Example: image compression with the SVD

1. **Big data and MapReduce**

* MapReduce: a framework for distributed computing
* Hadoop Streaming
* Running Hadoop jobs on Amazon Web Services
* Machine learning in MapReduce
* Using mrjob to automate MapReduce in Python
* Example: the Pegasos algorithm for distributed SVMs
* vector machines with mrjob
* Do you really need MapReduce?

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.



CATEGORY Data Analysis / BI

# Tika

Course Snapshot

* **Course:** Tika
* **Duration:** 4 days
* **Skill-level**: Foundation level Tika skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to learn how to pull usable information from otherwise inaccessible sources, including internet media and file archives.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Tika** is the ultimate guide to content mining using Apache Tika. You'll learn how to pull usable information from otherwise inaccessible sources, including internet media and file archives. This example-rich course teaches you to build and extend applications based on real-world experience with search engines, digital asset management, and scientific data processing. In addition to architectural overviews, you'll find detailed lessons on features like metadata extraction, automatic language detection, and custom parser development.

Working in a hands-on learning environment, led by our Tika expert instructor, students will learn about and explore:

* examples and case studies offer real-world experience from domains ranging from search engines to digital asset management and scientific data processing.
* You'll learn how to pull usable information from otherwise inaccessible sources, including internet media and file archives

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Crack MS Word, PDF, HTML, and ZIP
* Integrate with search engines, CMS, and other data sources
* Learn through experimentation
* Many examples

Audience & Pre-Requisites

This course is written for developers familiar with Java.

**Pre-Requisites:** Students should have

* requires no previous knowledge of Tika or text mining techniques.
* It assumes a working knowledge of Java.

Course Agenda / Topics

1. **The case for the digital Babel fishfree**

* Understanding digital documents
* everything
* What is Apache Tika?

1. **Getting started with Tika**

* Working with Tika source code
* The Tika application
* Tika as an embedded library

1. **The information landscape**

* Measuring information overload
* I’m feeling lucky—searching the information landscape
* Beyond lucky: machine learning

1. **Document type detection**

* Internet media types
* Media types in Tika
* File format diagnostics
* Tika, the type inspector

1. **Content extraction**

* Full-text extraction
* The Parser interface
* Document input stream
* Structured XHTML output
* Context-sensitive parsing

1. **Understanding metadata**

* The standards of metadata
* Metadata quality
* Metadata in Tika
* Practical uses of metadata

1. **Language detection**

* The most translated document in the world
* Sounds Greek to me—theory of language detection
* Language detection in Tika

1. **What���s in a file?free**

* Types of content
* How Tika extracts content

1. **The big picture**

* Tika in search engines
* Managing and mining information
* Buzzword compliance

1. **Tika and the Lucene search stack**

* Load-bearing walls
* The steel frame
* The finishing touches

1. **Extending Tika**

* Adding type information
* Custom type detection
* Customized parsing

1. **Powering NASA science data systems**

* NASA’s Planetary Data System
* NASA’s Earth Science Enterprise

1. **Content management with Apache Jackrabbit**

* Introducing Apache Jackrabbit
* The text extraction pool
* Content-aware WebDAV

1. **Curating cancer research data with Tika**

* The NCI Early Detection Research Network
* Integrating Tika

1. **The classic search engine example**

* The Public Terabyte Dataset Project
* The Bixo web crawler

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Pandas

**Course Snapshot**

* **Course:** Pandas
* **Duration:** 2 days
* **Skill-level**: Foundation-level Pandas skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to use pandas to automate repetitive spreadsheet functionality and derive insight from data by sorting columns, filtering data subsets, and creating multi-leveled indices.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Pandas makes it easy to dive into Python-based data analysis. You’ll learn to use pandas to automate repetitive spreadsheet functionality and derive insight from data by sorting columns, filtering data subsets, and creating multi-leveled indices. Each lesson is a self-contained tutorial, letting you dip in when you need to troubleshoot tricky problems. Best of all, you won’t be learning from sterile or randomly created data. You’ll start with a variety of datasets that are big, small, incomplete, broken, and messy and learn how to clean and format them for proper analysis

Working in a hands-on learning environment, led by our Pandas expert instructor, students will learn about and explore:

* Import a CSV, identify issues with its data structures, and convert it to the proper format
* Sort, filter, pivot, and draw conclusions from a dataset and its subsets.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Identify trends from text-based and time-based data
* Organize, group, merge, and join separate datasets
* Real-world datasets that are easy to download and explore

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to use pandas to automate repetitive spreadsheet functionality and derive insight from data by sorting columns, filtering data subsets, and creating multi-leveled indices

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Introducing Pandas**

* Data in the 21st Century
* Introducing pandas
* Importing a Dataset
* Manipulating a DataFrame
* Counting Values in a Series
* Filtering a Column by One or More Criteria
* Grouping Data

1. **Python Crash Course**

* Simple Data Types
* Operators
* Variables
* Functions
* Objects and Methods
* Lists
* Tuples
* Dictionaries
* Sets
* Modules, Classes, and Datetimes

1. **NumPy Crash Course**

* Dimensions
* The ndarray Object
* The nan Object

1. **The Series Object**

* Overview of a Series
* Create a Series from Python Objects
* Retrieving the First and Last Rows
* Mathematical Operations
* Passing the Series to Python's Built-In Functions
* Coding Challenges / Exercises

1. **Series Methods**

* Importing a Dataset with the read\_csv Method
* Sorting a Series
* Overwriting a Series with the inplace Parameter
* Counting Values with the value\_counts Method
* Invoking a Function on Every Series Value with the apply Method
* Coding Challenge: Deriving Insights from a Series

1. **The DataFrame Object**

* Overview of a DataFrame
* Similarities between Series and DataFrames
* Sorting a DataFrame
* Sort by Index
* Setting a New Index
* Selecting Columns or Rows from a DataFrame
* Select Rows from a DataFrame
* Extract Value from Series
* Rename Column or Row
* Resetting an Index
* Coding Challenge

1. **Filtering a DataFrame**

* Optimizing A Dataset for Memory Usage
* Filtering by a Single Condition
* Filtering by Multiple Conditions
* Filtering by Condition
* Dealing with Duplicates
* Coding Challenge

1. **Working with Text Data**

* String Casing
* String Slicing
* Boolean Methods
* Splitting Strings
* Coding Challenge
* A Note on Regular Expressions

1. **MultiIndex DataFrames**

* The MultiIndex Object
* MultiIndex DataFrames
* Sorting A MultiIndex
* Indexing with a MultiIndex
* Cross Sections
* Manipulating the Index

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Learning pandas

**Course Snapshot**

* **Course:** Learning pandas
* **Duration:** 4 days
* **Skill-level**: Foundation-level Pandas skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to get to grips with pandas—a versatile and high-performance Python library for data manipulation, analysis, and discovery
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

You will learn how to use pandas to perform data analysis in Python. You will start with an overview of data analysis and iteratively progress from modeling data, to accessing data from remote sources, performing numeric and statistical analysis, through indexing and performing aggregate analysis, and finally to visualizing statistical data and applying pandas to finance. With the knowledge you gain from this course, you will quickly learn pandas and how it can empower you in the exciting world of data manipulation, analysis and science.

Working in a hands-on learning environment, led by our Pandas expert instructor, students will learn about and explore:

* Get comfortable using pandas and Python as an effective data exploration and analysis tool
* Explore pandas through a framework of data analysis, with an explanation of how pandas is well suited for the various stages in a data analysis process
* A comprehensive guide to pandas with many of clear and practical examples to help you get up and using pandas

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand how data analysts and scientists think about of the processes of gathering and understanding data
* Learn how pandas can be used to support the end-to-end process of data analysis
* Use pandas Series and DataFrame objects to represent single and multivariate data
* Slicing and dicing data with pandas, as well as combining, grouping, and aggregating data from multiple sources
* How to access data from external sources such as files, databases, and web services
* Represent and manipulate time-series data and the many of the intricacies involved with this type of data
* How to visualize statistical information
* How to use pandas to solve several common data representation and analysis problems within finance

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to get to grips with pandas—a versatile and high-performance Python library for data manipulation, analysis, and discovery

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **pandas and Data Analysis**

* pandas and Data Analysis
* Introducing pandas
* Data manipulation, analysis, science, and pandas
* The process of data analysis
* Relating the book to the process
* Concepts of data and analysis in our tour of pandas
* Other Python libraries of value with pandas

1. **Up and Running with pandas**

* Up and Running with pandas
* Installation of Anaconda
* IPython and Jupyter Notebook
* Introducing the pandas Series and DataFrame
* Visualization

1. **Representing Univariate Data with the Series**

* Representing Univariate Data with the Series
* Configuring pandas
* Creating a Series
* The .index and .values properties
* The size and shape of a Series
* Specifying an index at creation
* Heads, tails, and takes
* Retrieving values in a Series by label or position
* Slicing a Series into subsets
* Alignment via index labels
* Performing Boolean selection
* Re-indexing a Series
* Modifying a Series in-place

1. **Representing Tabular and Multivariate Data with the DataFrame**

* Representing Tabular and Multivariate Data with the DataFrame
* Configuring pandas
* Creating DataFrame objects
* Accessing data within a DataFrame
* Selecting rows using Boolean selection
* Selecting across both rows and columns

1. **Manipulating DataFrame Structure**

* Manipulating DataFrame Structure
* Configuring pandas
* Renaming columns
* Adding new columns with [] and .insert()
* Adding columns through enlargement
* Adding columns using concatenation
* Reordering columns
* Replacing the contents of a column
* Deleting columns
* Appending new rows
* Concatenating rows
* Adding and replacing rows via enlargement
* Removing rows using .drop()
* Removing rows using Boolean selection
* Removing rows using a slice

1. **Indexing Data**

* Indexing Data
* Configuring pandas
* The importance of indexes
* The pandas index types
* Working with Indexes
* Hierarchical indexing

1. **Categorical Data**

* Categorical Data
* Configuring pandas
* Creating Categoricals
* Renaming categories
* Appending new categories
* Removing categories
* Removing unused categories
* Setting categories
* Descriptive information of a Categorical
* Munging school grades

1. **Numerical and Statistical Methods**

* Numerical and Statistical Methods
* Configuring pandas
* Performing numerical methods on pandas objects
* Performing statistical processes on pandas objects

1. **Accessing Data**

* Accessing Data
* Configuring pandas
* Working with CSV and text/tabular format data
* Reading and writing data in Excel format
* Reading and writing JSON files
* Reading HTML data from the web
* Reading and writing HDF5 format files
* Accessing CSV data on the web
* Reading and writing from/to SQL databases
* Reading data from remote data services

1. **Tidying Up Your Data**

* Tidying Up Your Data
* Configuring pandas
* What is tidying your data?
* How to work with missing data
* Handling duplicate data
* Transforming data

1. **Combining, Relating, and Reshaping Data**

* Combining, Relating, and Reshaping Data
* Configuring pandas
* Concatenating data in multiple objects
* Merging and joining data
* Pivoting data to and from value and indexes
* Stacking and unstacking
* Performance benefits of stacked data

1. **Data Aggregation**

* Data Aggregation
* Configuring pandas
* The split, apply, and combine (SAC) pattern
* Data for the examples
* Splitting data
* Applying aggregate functions, transforms, and filters
* Transforming groups of data
* Filtering groups from aggregation

1. **Time-Series Modelling**

* Time-Series Modelling
* Setting up the IPython notebook
* Representation of dates, time, and intervals
* Introducing time-series data
* Calculating new dates using offsets
* Representing durations of time using Period
* Handling holidays using calendars
* Normalizing timestamps using time zones
* Manipulating time-series data
* Time-series moving-window operations

1. **Visualization**

* Visualization
* Configuring pandas
* Plotting basics with pandas
* Creating time-series charts
* Common plots used in statistical analyses
* Manually rendering multiple plots in a single chart

1. **Historical Stock Price Analysis**

* Historical Stock Price Analysis
* Setting up the IPython notebook
* Obtaining and organizing stock data from Google
* Plotting time-series prices
* Plotting volume-series data
* Calculating the simple daily percentage change in closing price
* Calculating simple daily cumulative returns of a stock
* Resampling data from daily to monthly returns
* Analyzing distribution of returns
* Performing a moving-average calculation
* Comparison of average daily returns across stocks
* Correlation of stocks based on the daily percentage change of the closing price
* Calculating the volatility of stocks
* Determining risk relative to expected returns

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Analysis with NumPy and Pandas

**Course Snapshot**

* **Course:** Data Analysis with NumPy and Pandas
* **Duration:** 2 days
* **Skill-level**: Foundation-level Data Analysis with NumPy and Pandas skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Get to grips with the most popular Python packages that make Data Analysis possible.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Python, a multi-paradigm programming language, has become the language of choice for data scientists for visualization, data analysis, and machine learning. Data Analysis with NumPy and Pandas starts by guiding you in setting up the right environment for data analysis with Python, along with helping you install the correct Python distribution. In addition to this, you will work with the Jupyter notebook and set up a database. Once you have covered Jupyter, you will dig deep into Python’s NumPy package, a powerful extension with advanced mathematical functions. You will then move on to creating NumPy arrays and employing different array methods and functions. You will explore Python’s pandas extension which will help you get to grips with data mining and learn to subset your data. Last but not the least you will grasp how to manage your datasets by sorting and ranking them. By the end of this course, you will have learned to index and group your data for sophisticated data analysis and manipulation.

Working in a hands-on learning environment, led by our Data Analysis with NumPy and Pandas expert instructor, students will learn about and explore:

* Explore the tools you need to become a data analyst
* Discover practical examples to help you grasp data processing concepts
* Walk through hierarchical indexing and grouping for data analysis

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand how to install and manage Anaconda
* Read, sort, and map data using NumPy and pandas
* Find out how to create and slice data arrays using NumPy
* Discover how to subset your DataFrames using pandas
* Handle missing data in a pandas DataFrame
* Explore hierarchical indexing and plotting with pandas

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to get to grips with the most popular Python packages that make Data Analysis possible

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **Setting Up a Python Data Analysis Environment**

* Setting Up a Python Data Analysis Environment
* What is Anaconda?
* Installing Anaconda
* Exploring Jupyter Notebooks
* Exploring alternatives to Jupyter
* Package management with Conda
* Setting up a database

1. **Diving into NumPY**

* Diving into NumPY
* NumPy arrays
* Special numeric values
* Creating NumPy arrays

1. **Operations on NumPy Arrays**

* Operations on NumPy Arrays
* Selecting elements explicitly
* Advanced indexing
* Expanding arrays
* Arithmetic and linear algebra with arrays
* Linear algebra
* Employing array methods and functions

1. **pandas are Fun! What** **is** **pandas?**

* pandas are Fun! What is pandas?
* What does pandas do?
* Exploring series and DataFrame objects
* Subsetting your data
* Indexing methods

1. **Arithmetic, Function Application, and Mapping with pandas**

* Arithmetic, Function Application, and Mapping with pandas
* Arithmetic
* Handling missing data in a pandas DataFrame

1. **Managing, Indexing, and Plotting**

* Managing, Indexing, and Plotting
* Index sorting
* Hierarchical indexing
* Plotting with pandas

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Turning Spreadsheets into Corporate Data

**Course Snapshot**

* **Course:** Turning Spreadsheets into Corporate Data
* **Duration:** 4 days
* **Skill-level**: Foundation-level Spreadsheets skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Master the spreadsheet disambiguation technique and create credible and well-organized spreadsheets
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Spreadsheets are a popular way to store and communicate business data, but, although they are easy to create and update, they are not reliable enough to be used for making important corporate decisions. With this course, you can gain insight into how to maintain spreadsheets, how to format them, and then convert them into a database of reliable and useful information. Turning Spreadsheets into Corporate Data starts with a quick history of spreadsheet usage. You’ll learn the basics of formatting spreadsheets, including how to handle special characters and column headings, and how to convert the spreadsheet first into an intermediate database and then into corporate data. You will also learn how to utilize the mnemonic dictionary that is created along with the intermediate database. The later chapters discuss the immutability of data and the importance of organizational and political considerations during the data transformation. By the end of this course, you’ll have the skills and knowledge needed to convert your spreadsheets into reliable corporate data.

Working in a hands-on learning environment, led by our Spreadsheets expert instructor, students will learn about and explore:

* Gain insight into the different factors that affect the transformation of spreadsheets into data
* Explore in detail the basics of spreadsheet formatting
* Discover ways to handle non-standard spreadsheet structures

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Study the two ways of viewing spreadsheets: internal and external
* Establish guidelines for the end-user while submitting spreadsheets
* Discover ways to compensate for the shortcomings of the .pdf formatting
* Learn how to convert intermediate data from spreadsheets into corporate data
* Work with the mnemonic dictionary
* Understand data modeling through the entity-relationship diagram

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Master the spreadsheet disambiguation technique and create credible and well-organized spreadsheets

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **Brief History of Spreadsheets**

* The IT Labyrinth
* End User Acceptance of the Spreadsheet
* Spreadsheet Hell
* A Tradeoff
* Responsibility—The Flip Side of Control
* Management’s Problem
* Differences Between Two Types of Data

1. **Spreadsheet Paradox**

* Public Data
* The Spreadsheet as a Medium of Exchange
* Recurring/Non-Recurring Spreadsheets
* The Spectrum of Spreadsheets
* The Cost of Transforming a Spreadsheet
* Factors Other than Cost
* Transcription of Data
* Cell Formula
* Spreadsheet Descriptors
* Artificially Supplying Descriptors

1. **Spreadsheet Varieties**

* Simple Demarcation—xlstab
* Other Special Characters—eold and Linefeed
* The Internal View of a Spreadsheet
* A Missing Column Heading
* A Missing Value
* A Multiline Row
* The “Standard” Spreadsheet Format
* Managing the User of the Spreadsheet
* The ssdef Table
* The Spreadsheet Processing Log
* The Lineage of Spreadsheet Data
* The Cell Formula
* Relating to the Real World
* Identifying the Header Line

1. **The PDF Spreadsheet**

* The Importance of Special Characters
* PDF and OCR
* A Final Option

1. **The Basics of Spreadsheet Formatting**

* The System Name
* Unreliability of Report Name
* Multiple Sheets in a Spreadsheet
* Other Special Characters
* Identifying Column Headings
* Similar Column Headings
* Blocking Off Sections of a Spreadsheet
* Non-Standard Spreadsheet Structures
* A Spreadsheet that Cannot be Mapped
* A Spreadsheet in a TXT Format

1. **Spreadsheet Disambiguation**

* Selecting Spreadsheets for Inclusion into Corporate Data
* Recasting the Spreadsheet
* Logging the Spreadsheet for Transformation
* Entry into the Path Queue
* Defining the Spreadsheet Headings
* Pairing the ssdef Specification to the Spreadsheet
* Finding and Creating Database Definitions and Values
* The Intermediate Database
* Some Anomalies
* What if an Error is Discovered?
* Manual Effort Required
* Spreadsheet Width
* Subdividing a Spreadsheet
* No Value for a Column Name
* No Column Headings
* Creating the ssdef Specification Once

1. **The Intermediate Database**

* Finding Errors
* The Contents of the Intermediate Database
* Functions Served by the Data Elements
* Alternate Name
* Adding Context to Data Values
* Editing Data in the Intermediate Database

1. **The ssdef Database**

* Organizing Data Inside the ssdef Table
* Processing Using ssdef Records
* Searching the Full Path Queue

1. **The Corporate Database**

* From Intermediate Data to Corporate Data
* Grouped Corporate Data
* Tracing the Lineage

1. **The Mnemonic Dictionary**

* The Contents of the Mnemonic Dictionary
* Grouping Like Data Elements
* Applying Naming Conventions
* Value of the Mnemonic Dictionary

1. **Political Considerations Within the Organization**

* Shifting Control
* Immutability of Data
* The Importance of Alternate Names
* Limited Editing
* Super Classifications of Data
* The Lineage of Corporate Data
* Relative Volumes of Data

1. **Data Modeling and the Spreadsheet Environment**

* The Entity Relationship Diagram
* The Data Item Set
* The Physical Model
* The Data Model
* The Data Model and Spreadsheet Data
* “Correctness” of Data
* Aligning Data from Different Spreadsheets
* An Algorithmic Resolution
* An Indexed Resolution
* Resolution and the Data Model
* Spreadsheet Data in the Data Warehouse
* Changing Spreadsheet Data

1. **Case Study**

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Learning Jupyter 5

**Course Snapshot**

* **Course:** Learning Jupyter 5
* **Duration:** 3 days
* **Skill-level**: Foundation-level Jupyter skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Create and share livecode, equations, visualizations, and explanatory text, in both a single document and a web browser with Jupyter.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The Jupyter Notebook allows you to create and share documents that contain live code, equations, visualizations, and explanatory text. The Jupyter Notebook system is extensively used in domains such as data cleaning and transformation, numerical simulation, statistical modeling, and machine learning. Learning Jupyter 5 will help you get to grips with interactive computing using real-world examples. The bo course ok starts with a detailed overview of the Jupyter Notebook system and its installation in different environments. Next, you will learn to integrate the Jupyter system with different programming languages such as R, Python, Java, JavaScript, and Julia, and explore various versions and packages that are compatible with the Notebook system. Moving ahead, you will master interactive widgets and namespaces and work with Jupyter in a multi-user mode. By the end of this course, you will have used Jupyter with a big dataset and be able to apply all the functionalities you’ve explored throughout the course. You will also have learned all about the Jupyter Notebook and be able to start performing data transformation, numerical simulation, and data visualization.

Working in a hands-on learning environment, led by our Jupyter expert instructor, students will learn about and explore:

* Learn how to use Jupyter 5.x features such as cell tagging and attractive table styles
* Leverage big data tools and datasets with different Python packages
* Explore multiple-user Jupyter Notebook servers

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Install and run the Jupyter Notebook system on your machine
* Implement programming languages such as R, Python, Julia, and JavaScript with the Jupyter Notebook
* Use interactive widgets to manipulate and visualize data in real time
* Start sharing your Notebook with colleagues
* Invite your colleagues to work with you on the same Notebook
* Organize your Notebook using Jupyter namespaces
* Access big data in Jupyter for dealing with large datasets using Spark

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Create and share livecode, equations, visualizations, and explanatory text, in both a single document and a web browser with Jupyter

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **Introduction to Jupyter**

* Introduction to Jupyter
* First look at Jupyter
* Installing Jupyter
* Notebook structure
* Notebook workflow
* Basic Notebook operations
* Security in Jupyter
* Configuration options for Jupyter

1. **Jupyter Python Scripting**

* Jupyter Python Scripting
* Basic Python in Jupyter
* Python data access in Jupyter
* Python pandas in Jupyter
* Python graphics in Jupyter
* Python random numbers in Jupyter

1. **Jupyter R Scripting**

* Jupyter R Scripting
* Adding R scripting to your installation
* Basic R in Jupyter
* R dataset access
* R visualizations in Jupyter
* R cluster analysis
* R forecasting
* R machine learning

1. **Jupyter Julia Scripting**

* Jupyter Julia Scripting
* Adding Julia scripting to your installation
* Basic Julia in Jupyter
* Julia limitations in Jupyter
* Standard Julia capabilities
* Julia visualizations in Jupyter
* Julia Vega plotting
* Julia parallel processing
* Julia control flow
* Julia regular expressions
* Julia unit testing

1. **Jupyter Java Coding**

* Jupyter Java Coding
* Adding the Java kernel to your installation
* Jupyter Java console
* Jupyter Java output
* Java Optional
* Java compiler errors
* Java lambdas
* Java Collections
* Java summary statistics

1. **Jupyter JavaScript Coding**

* Jupyter JavaScript Coding
* Adding JavaScript scripting to your installation
* JavaScript Hello World Jupyter Notebook
* Basic JavaScript in Jupyter
* JavaScript limitations in Jupyter
* Node.js d3 package
* Node.js stats-analysis package
* Node.js JSON handling
* Node.js canvas package
* Node.js plotly package
* Node.js asynchronous threads
* Node.js decision-tree package

1. **Jupyter Scala**

* Jupyter Scala
* Installing the Scala kernel
* Scala data access in Jupyter
* Scala array operations
* Scala random numbers in Jupyter
* Scala closures
* Scala higher-order functions
* Scala pattern matching
* Scala case classes
* Scala immutability
* Scala collections
* Named arguments
* Scala traits

1. **Jupyter and Big Data**

* Jupyter and Big Data
* Apache Spark
* First Spark script
* Spark word count
* Sorted word count
* Estimate pi
* Log file examination
* Spark primes
* Spark text file analysis
* Spark evaluating history data

1. **Interactive Widgets**

* Interactive Widgets
* Installing widgets
* Widget basics
* Interact widget
* Interactive widget
* Widgets

1. **Sharing and Converting Jupyter Notebooks**

* Sharing and Converting Jupyter Notebooks
* Sharing Notebooks
* Converting Notebooks

1. **Multiuser Jupyter Notebooks**

* Multiuser Jupyter Notebooks
* A sample interactive Notebook
* JupyterHub
* Docker

1. **What's Next?**

* What's Next?
* JupyterHub
* JupyterLab
* Scale
* Custom frontends
* Interactive computing standards

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Wrangling with Python

**Course Snapshot**

* **Course:** Data Wrangling with Python
* **Duration:** 2 days
* **Skill-level**: Foundation-level Data Wrangling with Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Simplify your ETL processes with these hands-on data hygiene tips, tricks, and best practices.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

For data to be useful and meaningful, it must be curated and refined. Data Wrangling with Python teaches you the core ideas behind these processes and equips you with knowledge of the most popular tools and techniques in the domain. The course starts with the absolute basics of Python, focusing mainly on data structures. It then delves into the fundamental tools of data wrangling like NumPy and Pandas libraries. You’ll explore useful insights into why you should stay away from traditional ways of data cleaning, as done in other languages, and take advantage of the specialized pre-built routines in Python. This combination of Python tips and tricks will also demonstrate how to use the same Python backend and extract/transform data from an array of sources including the Internet, large database vaults, and Excel financial tables. To help you prepare for more challenging scenarios, you’ll cover how to handle missing or wrong data, and reformat it based on the requirements from the downstream analytics tool. The course will further help you grasp concepts through real-world examples and datasets. By the end of this course, you will be confident in using a diverse array of sources to extract, clean, transform, and format your data efficiently.

Working in a hands-on learning environment, led by our Data Wrangling with Python expert instructor, students will learn about and explore:

* Focus on the basics of data wrangling
* Study various ways to extract the most out of your data in less time
* Boost your learning curve with bonus topics like random data generation and data integrity checks

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use and manipulate complex and simple data structures
* Harness the full potential of DataFrames and numpy.array at run time
* Perform web scraping with BeautifulSoup4 and html5lib
* Execute advanced string search and manipulation with RegEX
* Handle outliers and perform data imputation with Pandas
* Use descriptive statistics and plotting techniques
* Practice data wrangling and modeling using data generation techniques

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Simplify your ETL processes with these hands-on data hygiene tips, tricks, and best practices.

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Introduction to Data Wrangling with Python**

* Introduction to Data Wrangling with Python
* Introduction
* Python for Data Wrangling
* Lists, Sets, Strings, Tuples, and Dictionaries

1. **Advanced Data Structures and File Handling**

* Advanced Data Structures and File Handling
* Introduction
* Advanced Data Structures
* Basic File Operations in Python

1. **Introduction to NumPy, Pandas, and Matplotlib**

* Introduction to NumPy, Pandas, and Matplotlib
* Introduction
* NumPy Arrays
* Pandas DataFrames
* Statistics and Visualization with NumPy and Pandas

1. **A Deep Dive into Data Wrangling with Python**

* A Deep Dive into Data Wrangling with Python
* Introduction
* Subsetting, Filtering, and Grouping
* Detecting Outliers and Handling Missing Values
* Concatenating, Merging, and Joining
* Useful Methods of Pandas

1. **Getting Comfortable with** **Different Kinds of Data Sources**

* Getting Comfortable with Different Kinds of Data Sources
* Introduction
* Reading Data from Different Text-Based (and Non-Text-Based) Sources
* Introduction to Beautiful Soup 4 and Web Page Parsing

1. **Learning the Hidden Secrets of Data Wrangling**

* Learning the Hidden Secrets of Data Wrangling
* Introduction
* Advanced List Comprehension and the zip Function
* Data Formatting
* Identify and Clean Outliers
* Activity 8: Handling Outliers and Missing Data

1. **Advanced Web Scraping and Data Gathering**

* Advanced Web Scraping and Data Gathering
* Introduction
* The Basics of Web Scraping and the Beautiful Soup Library
* Reading Data from XML
* Reading Data from an API
* Fundamentals of Regular Expressions (RegEx)

1. **RDBMS and SQL**

* RDBMS and SQL
* Introduction
* Refresher of RDBMS and SQL
* Using an RDBMS (MySQL/PostgreSQL/SQLite)

1. **Application of Data Wrangling in Real Life**

* Application of Data Wrangling in Real Life
* Introduction
* Applying Your Knowledge to a Real-life Data Wrangling Task
* Activity 12: Data Wrangling Task – Fixing UN Data
* Activity 13: Data Wrangling Task – Cleaning GDP Data
* Activity 14: Data Wrangling Task – Merging UN Data and GDP Data
* Activity 15: Data Wrangling Task – Connecting the New Data to the Database
* An Extension to Data Wrangling

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Practical Data Wrangling

**Course Snapshot**

* **Course:** Practical Data Wrangling
* **Duration:** 2 days
* **Skill-level**: Foundation-level Data Wrangling skills for Intermediate skilled team members. This is not a basic class.
* **Target**ed Audience: This course is geared for Python experienced developers, analysts or others who wants to turn your noisy data into relevant, insight-ready information by leveraging the data wrangling techniques in Python and R
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Around 80% of time in data analysis is spent on cleaning and preparing data for analysis. This is, however, an important task, and is a prerequisite to the rest of the data analysis workflow, including visualization, analysis and reporting. Python and R are considered a popular choice of tool for data analysis, and have packages that can be best used to manipulate different kinds of data, as per your requirements. This course will show you the different data wrangling techniques, and how you can leverage the power of Python and R packages to implement them. You’ll start by understanding the data wrangling process and get a solid foundation to work with different types of data. You’ll work with different data structures and acquire and parse data from various locations. You’ll also see how to reshape the layout of data and manipulate, summarize, and join data sets. Finally, we conclude with a quick primer on accessing and processing data from databases, conducting data exploration, and storing and retrieving data quickly using databases. The course includes practical examples on each of these points using simple and real-world data sets to give you an easier understanding. By the end of the course, you’ll have a thorough understanding of all the data wrangling concepts and how to implement them in the best possible way.

Working in a hands-on learning environment, led by our Data Wrangling expert instructor, students will learn about and explore:

* This easy-to-follow guide takes you through every step of the data wrangling process in the best possible way
* Work with different types of datasets, and reshape the layout of your data to make it easier for analysis
* Get simple examples and real-life data wrangling solutions for data pre-processing

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Read a csv file into python and R, and print out some statistics on the data
* Gain knowledge of the data formats and programming structures involved in retrieving API data
* Make effective use of regular expressions in the data wrangling process
* Explore the tools and packages available to prepare numerical data for analysis
* Find out how to have better control over manipulating the structure of the data
* Create a dexterity to programmatically read, audit, correct, and shape data
* Write and complete programs to take in, format, and output data sets

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to turn your noisy data into relevant, insight-ready information by leveraging the data wrangling techniques in Python and R

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Programming with Data**

* Programming with Data
* Understanding data wrangling
* The tools for data wrangling

1. **Introduction to Programming in Python**

* Introduction to Programming in Python
* External resources
* Logistical overview
* Running programs in python
* Data types, variables, and the Python shell
* Compound statements
* Making annotations within programs
* A programmer's resources

1. **Reading, Exploring, and Modifying Data - Part I**

* Reading, Exploring, and Modifying Data - Part I
* External resources
* Logistical overview
* Introducing a basic data wrangling work flow
* Introducing the JSON file format
* Opening and closing a file in Python using file I/O
* Reading the contents of a file
* Exploring the contents of a data file
* Modifying a dataset
* Outputting the modified data to a new file
* Specifying input and output file names in the Terminal

1. **Reading, Exploring, and Modifying Data - Part II**

* Reading, Exploring, and Modifying Data - Part II
* Logistical overview
* Understanding the CSV format
* Introducing the CSV module
* Using the CSV module to read CSV data
* Using the CSV module to write CSV data
* Using the pandas module to read and process data
* Handling non-standard CSV encoding and dialect
* Understanding XML
* Using the XML module to parse XML data

1. **Manipulating Text Data - An Introduction to Regular Expressions**

* Manipulating Text Data - An Introduction to Regular Expressions
* Logistical overview
* Understanding the need for pattern recognition
* Introducting regular expressions
* Looking for patterns
* Quantifying the existence of patterns
* Extracting patterns
* Summary

1. **6Cleaning Numerical Data - An Introduction to R and RStudio**

* Cleaning Numerical Data - An Introduction to R and RStudio
* Logistical overview
* Introducing R and RStudio
* Familiarizing yourself with RStudio
* Conducting basic outlier detection and removal
* Handling NA values
* Variable names and contents

1. **Simplifying Data Manipulation with dplyr**

* Simplifying Data Manipulation with dplyr
* Logistical overview
* Introducing dplyr
* Getting started with dplyr
* Chaining operations together
* Filtering the rows of a dataframe
* Summarizing data by category
* Rewriting code using dplyr

1. **Getting Data from the Web**

* Getting Data from the Web
* Logistical overview
* Introducing APIs
* Using Python to retrieve data from APIs
* Using URL parameters to filter the results

1. **Working with Large Datasets**

* Working with Large Datasets
* Logistical overview
* Understanding computer memory
* Understanding databases
* Introducing MongoDB
* Interfacing with MongoDB from Python

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Python Data Structures and Algorithms

**Course Snapshot**

* **Course:** Python Data Structures and Algorithms
* **Duration:** 3 days
* **Skill-level**: Foundation-level Python Data Structures and Algorithms skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to implement classic and functional data structures and algorithms using Python
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Data structures allow you to organize data in a particular way efficiently. They are critical to any problem, provide a complete solution, and act like reusable code. In this course, you will learn the essential Python data structures and the most common algorithms. With this easy-to-read book, you will be able to understand the power of linked lists, double linked lists, and circular linked lists. You will be able to create complex data structures such as graphs, stacks and queues. We will explore the application of binary searches and binary search trees. You will learn the common techniques and structures used in tasks such as preprocessing, modeling, and transforming data. We will also discuss how to organize your code in a manageable, consistent, and extendable way. The course will explore in detail sorting algorithms such as bubble sort, selection sort, insertion sort, and merge sort. By the end of the course, you will learn how to build components that are easy to understand, debug, and use in different applications.

Working in a hands-on learning environment, led by our Python Data Structures and Algorithms expert instructor, students will learn about and explore:

* A step by step guide, which will provide you with a thorough discussion on the analysis and design of fundamental Python data structures.
* Get a better understanding of advanced Python concepts such as big-o notation, dynamic programming, and functional data structures.
* Explore illustrations to present data structures and algorithms, as well as their analysis, in a clear, visual manner.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Gain a solid understanding of Python data structures.
* Build sophisticated data applications.
* Understand the common programming patterns and algorithms used in Python data science.
* Write efficient robust code.

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Implement classic and functional data structures and algorithms using Python.

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Python Objects, Types, and Expressions**

* Python Objects, Types, and Expressions
* Understanding data structures and algorithms
* Python for data

1. **Python Data Types and Structures**

* Python Data Types and Structures
* Operations and expressions
* Built-in data types
* Sets
* Modules for data structures and algorithms

1. **Principles of Algorithm Design**

* Principles of Algorithm Design
* Algorithm design paradigms
* Recursion and backtracking
* Runtime analysis
* Amortized analysis

1. **Lists and Pointer Structures**

* Lists and Pointer Structures
* Arrays
* Pointer structures
* Nodes
* Finding endpoints
* Singly linked lists
* A faster append operation
* Getting the size of the list
* Improving list traversal
* Deleting nodes
* Clearing a list
* Doubly linked lists
* Circular lists

1. **Stacks and Queues**

* Stacks and Queues
* Stacks
* Queues

1. **Trees**

* Trees
* Terminology
* Tree nodes
* Binary trees

1. **Hashing and Symbol Tables**

* Hashing and Symbol Tables
* Hashing
* Hash table

1. **Graphs and Other Algorithms**

* Graphs and Other Algorithms
* Graphs
* Directed and undirected graphs
* Weighted graphs
* Graph representation
* Graph traversal
* Other useful graph methods
* Priority queues and heaps
* Selection algorithms

1. **Searching**

* Searching
* Linear Search
* Binary search
* Interpolation search

1. **Sorting**

* Sorting
* Sorting algorithms
* Bubble sort
* Insertion sort
* Selection sort
* Quick sort

1. **Selection Algorithms**

* Selection Algorithms
* Selection by sorting
* Randomized selection
* Deterministic selection

1. **Design Techniques and Strategies**

* Design Techniques and Strategies
* Classification of algorithms
* Technical implementation
* Complexity classes

1. **Implementations, Applications, and Tools**

* Implementations, Applications, and Tools
* Tools of the trade
* Data preprocessing
* Machine learning
* Data visualization

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Analysis with Python

**Course Snapshot**

* **Course:** Data Analysis with Python
* **Duration:** 3 days
* **Skill-level**: Foundation-level Data Analysis with Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Learn a modern approach to data analysis using Python to harness the power of programming and AI across your data. Detailed case studies bring this modern approach to life across visual data, social media, graph algorithms, and time series analysis.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Data Analysis with Python offers a modern approach to data analysis so that you can work with the latest and most powerful Python tools, AI techniques, and open source libraries. Industry expert David Taieb shows you how to bridge data science with the power of programming and algorithms in Python. You'll be working with complex algorithms, and cutting-edge AI in your data analysis. Learn how to analyze data with hands-on examples using Python-based tools and Jupyter Notebook. You'll find the right balance of theory and practice, with extensive code files that you can integrate right into your own data projects. Explore the power of this approach to data analysis by then working with it across key industry case studies. Four fascinating and full projects connect you to the most critical data analysis challenges you’re likely to meet in today. The first of these is an image recognition application with TensorFlow – embracing the importance today of AI in your data analysis. The second industry project analyses social media trends, exploring big data issues and AI approaches to natural language processing. The third case study is a financial portfolio analysis application that engages you with time series analysis - pivotal to many data science applications today. The fourth industry use case dives you into graph algorithms and the power of programming in modern data science. You'll wrap up with a thoughtful look at the future of data science and how it will harness the power of algorithms and artificial intelligence.

Working in a hands-on learning environment, led by our Data Analysis with Python expert instructor, students will learn about and explore:

* Bridge your data analysis with the power of programming, complex algorithms, and AI
* Use Python and its extensive libraries to power your way to new levels of data insight
* Work with AI algorithms, TensorFlow, graph algorithms, NLP, and financial time series
* Explore this modern approach across with key industry case studies and hands-on projects

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* A new toolset that has been carefully crafted to meet for your data analysis challenges
* Full and detailed case studies of the toolset across several of today’s key industry contexts
* Become super productive with a new toolset across Python and Jupyter Notebook
* Look into the future of data science and which directions to develop your skills next

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Learn a modern approach to data analysis using Python to harness the power of programming and AI across your data. Detailed case studies bring this modern approach to life across visual data, social media, graph algorithms, and time series analysis.

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **Programming and Data Science – A New Toolset**

* Programming and Data Science – A New Toolset
* What is data science
* Is data science here to stay?
* Why is data science on the rise?
* What does that have to do with developers?
* Putting these concepts into practice
* Deep diving into a concrete example
* Data pipeline blueprint
* What kind of skills are required to become a data scientist?
* IBM Watson DeepQA
* Back to our sentiment analysis of Twitter hashtags project
* Lessons learned from building our first enterprise-ready data pipeline
* Data science strategy
* Jupyter Notebooks at the center of our strategy

1. **Python and Jupyter Notebooks to Power your Data Analysis**

* Python and Jupyter Notebooks to Power your Data Analysis
* Why choose Python?
* Introducing PixieDust
* SampleData – a simple API for loading data
* Wrangling data with pixiedust\_rosie
* Display – a simple interactive API for data visualization
* Filtering
* Bridging the gap between developers and data scientists with PixieApps
* Architecture for operationalizing data science analytics

1. **Accelerate your Data Analysis with Python Libraries**

* Accelerate your Data Analysis with Python Libraries
* Anatomy of a PixieApp

1. **Publish your Data Analysis** **to the Web - the PixieApp Tool**

* Publish your Data Analysis to the Web - the PixieApp Tool
* Overview of Kubernetes
* Installing and configuring the PixieGateway server

1. **Python and PixieDust Best Practices and Advanced Concepts**

* Python and PixieDust Best Practices and Advanced Concepts
* Use @captureOutput decorator to integrate the output of third-party Python libraries
* Increase modularity and code reuse
* Run Node.js inside a Python Notebook

1. **Analytics Study: AI and Image Recognition with TensorFlow**

* Analytics Study: AI and Image Recognition with TensorFlow
* What is machine learning?
* What is deep learning?
* Getting started with TensorFlow
* Image recognition sample application

1. **Analytics Study: NLP and Big Data with Twitter Sentiment Analysis**

* Analytics Study: NLP and Big Data with Twitter Sentiment Analysis
* Getting started with Apache Spark
* Twitter sentiment analysis application
* Part 1 – Acquiring the data with Spark Structured Streaming
* Part 2 – Enriching the data with sentiment and most relevant extracted entity
* Part 3 – Creating a real-time dashboard PixieApp
* Part 4 – Adding scalability with Apache Kafka and IBM Streams Designer

1. **Analytics Study: Prediction - Financial Time Series Analysis and Forecasting**

* Analytics Study: Prediction - Financial Time Series Analysis and Forecasting
* Getting started with NumPy
* Statistical exploration of time series
* Putting it all together with the StockExplorer PixieApp
* Time series forecasting using the ARIMA model

1. **Analytics Study: Graph Algorithms - US Domestic Flight Data Analysis**

* Analytics Study: Graph Algorithms - US Domestic Flight Data Analysis
* Introduction to graphs
* Getting started with the networkx graph library
* Part 1 – Loading the US domestic flight data into a graph
* Part 2 – Creating the USFlightsAnalysis PixieApp
* Part 3 – Adding data exploration to the USFlightsAnalysis PixieApp
* Part 4 – Creating an ARIMA model for predicting flight delays

1. **The Future of Data Analysis and Where to Develop your Skills**

* The Future of Data Analysis and Where to Develop your Skills
* Forward thinking – what to expect for AI and data science

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Go Web Scraping

**Course Snapshot**

* **Course:** Go Web Scraping
* **Duration:** 2 days
* **Skill-level**: Foundation-level Go Web Scraping skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Learn how some Go-specific language features help to simplify building web scrapers along with common pitfalls and best practices regarding web scraping.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Web scraping is the process of extracting information from the web using various tools that perform scraping and crawling. Go is emerging as the language of choice for scraping using a variety of libraries. This course will quickly explain to you, how to scrape data data from various websites using Go libraries such as Colly and Goquery. The course starts with an introduction to the use cases of building a web scraper and the main features of the Go programming language, along with setting up a Go environment. It then moves on to HTTP requests and responses and talks about how Go handles them. You will also learn about a number of basic web scraping etiquettes. You will be taught how to navigate through a website, using a breadth-first and then a depth-first search, as well as find and follow links. You will get to know about the ways to track history in order to avoid loops and to protect your web scraper using proxies. Finally the course will cover the Go concurrency model, and how to run scrapers in parallel, along with large-scale distributed web scraping.

Working in a hands-on learning environment, led by our Go Web Scraping expert instructor, students will learn about and explore:

* Use Go libraries like Goquery and Colly to scrape the web
* Common pitfalls and best practices to effectively scrape and crawl
* Learn how to scrape using the Go concurrency model

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Implement Cache-Control to avoid unnecessary network calls
* Coordinate concurrent scrapers
* Design a custom, larger-scale scraping system
* Scrape basic HTML pages with Colly and JavaScript pages with chromedp
* Discover how to search using the "strings" and "regexp" packages
* Set up a Go development environment
* Retrieve information from an HTML document
* Protect your web scraper from being blocked by using proxies
* Control web browsers to scrape JavaScript sites

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Learn how some Go-specific language features help to simplify building web scrapers along with common pitfalls and best practices regarding web scraping.

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **Introducing Web Scraping and Go**

* Introducing Web Scraping and Go
* What is web scraping?
* Why do you need a web scraper?
* What is Go?
* Why is Go a good fit for web scraping?
* How to set up a Go development environment

1. **The Request/Response Cycle**

* The Request/Response Cycle
* What do HTTP requests look like?
* What do HTTP responses look like?
* What are HTTP status codes?
* What do HTTP requests/responses look like in Go?

1. **Web Scraping Etiquette**

* Web Scraping Etiquette
* What is a robots.txt file?
* What is a User-Agent string?
* How to throttle your scraper
* How to use caching

1. **Parsing HTML**

* Parsing HTML
* What is the HTML format?
* Searching using the strings package
* Searching using the regexp package
* Searching using XPath queries
* Searching using Cascading Style Sheets selectors

1. **Web Scraping Navigation**

* Web Scraping Navigation
* Following links
* Submitting forms
* Avoiding loops
* Breadth-first versus depth-first crawling
* Navigating with JavaScript

1. **Protecting Your Web Scraper**

* Protecting Your Web Scraper
* Virtual private servers
* Proxies
* Virtual private networks
* Boundaries

1. **Scraping with Concurrency**

* Scraping with Concurrency
* What is concurrency
* Concurrency pitfalls
* The Go concurrency model
* sync package helpers

1. **Scraping at 100x**

* Scraping at 100x
* Components of a web scraping system
* Scraping HTML pages with colly
* Scraping JavaScript pages with chrome-protocol
* Distributed scraping with dataflowkit
* Summary

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Feature Engineering Made Easy

**Course Snapshot**

* **Course:** Feature Engineering Made Easy
* **Duration:** 2 days
* **Skill-level**: Foundation-level Feature Engineering skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for experienced developers, analysts or others who wants a perfect guide to speed up the predicting power of Machine Learning Algorithms
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Feature engineering is the most important step in creating powerful machine learning systems. This course will take you through the entire feature-engineering journey to make your machine learning much more systematic and effective. You will start with understanding your data—often the success of your ML models depends on how you leverage different feature types, such as continuous, categorical, and more, You will learn when to include a feature, when to omit it, and why, all by understanding error analysis and the acceptability of your models. You will learn to convert a problem statement into useful new features. You will learn to deliver features driven by business needs as well as mathematical insights. You'll also learn how to use machine learning on your machines, automatically learning amazing features for your data. By the end of the course, you will become proficient in Feature Selection, Feature Learning, and Feature Optimization.

Working in a hands-on learning environment, led by our Feature Engineering expert instructor, students will learn about and explore:

* Design, discover, and create dynamic, efficient features for your machine learning application
* Understand your data in-depth and derive astonishing data insights with the help of this Guide
* Grasp powerful feature-engineering techniques and build machine learning systems

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Identify and leverage different feature types
* Clean features in data to improve predictive power
* Understand why and how to perform feature selection, and model error analysis
* Leverage domain knowledge to construct new features
* Deliver features based on mathematical insights
* Use machine-learning algorithms to construct features
* Master feature engineering and optimization
* Harness feature engineering for real world applications through a structured case study

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wants a perfect guide to speed up the predicting power of Machine Learning Algorithms.

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **Introduction to Feature Engineering**

* Introduction to Feature Engineering
* Motivating example – AI-powered communications
* Why feature engineering matters
* What is feature engineering?
* Evaluation of machine learning algorithms and feature engineering procedures
* Feature understanding – what’s in my dataset?
* Feature improvement – cleaning datasets
* Feature selection – say no to bad attributes
* Feature construction – can we build it?
* Feature transformation – enter math-man
* Feature learning – using AI to better our AI

1. **Feature Understanding – What's in My Dataset?**

* Feature Understanding – What's in My Dataset?
* The structure, or lack thereof, of data
* An example of unstructured data – server logs
* Quantitative versus qualitative data
* The four levels of data
* Recap of the levels of data

1. **Feature Improvement - Cleaning Datasets**

* Feature Improvement - Cleaning Datasets
* Identifying missing values in data
* Dealing with missing values in a dataset
* Standardization and normalization

1. **Feature Construction**

* Feature Construction
* Examining our dataset
* Imputing categorical features
* Encoding categorical variables
* Extending numerical features
* Text-specific feature construction

1. **Feature Selection**

* Feature Selection
* Achieving better performance in feature engineering
* Creating a baseline machine learning pipeline
* The types of feature selection
* Choosing the right feature selection method

1. **Feature Transformations**

* Feature Transformations
* Dimension reduction – feature transformations versus feature selection versus feature construction
* Principal Component Analysis
* Scikit-learn's PCA
* How centering and scaling data affects PCA
* A deeper look into the principal components
* Linear Discriminant Analysis
* LDA versus PCA – iris dataset

1. **Feature Learning**

* Feature Learning
* Parametric assumptions of data
* Restricted Boltzmann Machines
* The BernoulliRBM
* Extracting RBM components from MNIST
* Using RBMs in a machine learning pipeline
* Learning text features – word vectorizations

1. **Case Studies**

* Case Studies
* Case study 1 - facial recognition
* Case study 2 - predicting topics of hotel reviews data

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Analysis with Scala

**Course Snapshot**

* **Course:** Data Analysis with Scala
* **Duration:** 2 days
* **Skill-level**: Foundation-level Data Analysis with Scala skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Master scala's advanced techniques to solve real-world problems in data analysis and gain valuable insights from your data.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Efficient business decisions with an accurate sense of business data helps in delivering better performance across products and services. This course helps you to leverage the popular Scala libraries and tools for performing core data analysis tasks with ease. The course begins with a quick overview of the building blocks of a standard data analysis process. You will learn to perform basic tasks like Extraction, Staging, Validation, Cleaning, and Shaping of datasets. You will later deep dive into the data exploration and visualization areas of the data analysis life cycle. You will make use of popular Scala libraries like Saddle, Breeze, Vegas, and Prediction for processing your datasets. You will learn statistical methods for deriving meaningful insights from data. You will also learn to create applications for Apache Spark 2.x on complex data analysis, in real-time. You will discover traditional machine learning techniques for doing data analysis. Furthermore, you will also be introduced to neural networks and deep learning from a data analysis standpoint. By the end of this course, you will be capable of handling large sets of structured and unstructured data, perform exploratory analysis, and building efficient Scala applications for discovering and delivering insights

Working in a hands-on learning environment, led by our Data Analysis with Scala expert instructor, students will learn about and explore:

* A beginner's guide for performing data analysis loaded with numerous rich, practical examples
* Access to popular Scala libraries such as Breeze, Saddle for efficient data manipulation and exploratory analysis
* Develop applications in Scala for real-time analysis and machine learning in Apache Spark.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Techniques to determine the validity and confidence level of data
* Apply quartiles and n-tiles to datasets to see how data is distributed into many buckets
* Create data pipelines that combine multiple data lifecycle steps
* Use built-in features to gain a deeper understanding of the data
* Apply Lasso regression analysis method to your data
* Compare Apache Spark API with traditional Apache Spark data analysis

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Master scala's advanced techniques to solve real-world problems in data analysis and gain valuable insights from your data

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Scala Overview**

* Scala Overview
* Getting started with Scala
* Overview of object-oriented and functional programming
* Scala case classes and the collection API
* Overview of Scala libraries for data analysis

1. **Data Analysis Life Cycle**

* Data Analysis Life Cycle
* Data journey
* Sourcing data
* Understanding data
* Using ML to learn from data
* Creating a data pipeline

1. **Data Ingestion**

* Data Ingestion
* Data extraction
* Data staging
* Cleaning and normalizing
* Enriching
* Organizing and storing

1. **Data Exploration and Visualization**

* Data Exploration and Visualization
* Sampling data
* Performing ad hoc analysis
* Finding a relationship between data elements
* Visualizing data

1. **Applying Statistics and Hypothesis Testing**

* Applying Statistics and Hypothesis Testing
* Basics of statistics
* Vector level statistics
* Random data generation
* Hypothesis testing

1. **Introduction to Spark for Distributed Data Analysis**

* Introduction to Spark for Distributed Data Analysis
* Spark setup and overview
* Spark Datasets and DataFrames
* Sourcing data using Spark
* Using Spark to explore data

1. **Traditional Machine Learning for Data Analysis**

* Traditional Machine Learning for Data Analysis
* ML overview
* Decision trees
* Random forest
* Ridge and lasso regression
* k-means cluster analysis
* Natural language processing for data analysis
* Algorithm selections

1. **Near Real-Time Data Analysis Using Streaming**

* Near Real-Time Data Analysis Using Streaming
* Overview of streaming
* Spark Streaming overview
* Streaming a k-means clustering algorithm using Spark
* Streaming linear regression using Spark

1. **Working with Data at Scale**

* Working with Data at Scale
* Working with data at scale
* Cost considerations
* Reliability considerations

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Learning Alteryx

**Course Snapshot**

* **Course:** Learning Alteryx
* **Duration:** 3 days
* **Skill-level**: Foundation-level Alteryx skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for experienced developers, analysts or others who wants to Implement a Business Intelligence solutions without any coding - by leveraging the power of the Alteryx platform
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Alteryx, as a leading data blending and advanced data analytics platform, has taken self-service data analytics to the next level. Companies worldwide often find themselves struggling to prepare and blend massive datasets that are time-consuming for analysts. Alteryx solves these problems with a repeatable workflow designed to quickly clean, prepare, blend, and join your data in a seamless manner. This course will set you on a self-service data analytics journey that will help you create efficient workflows using Alteryx, without any coding involved. It will empower you and your organization to take well-informed decisions with the help of deeper business insights from the data. Starting with the fundamentals of using Alteryx such as data preparation and blending, you will delve into the more advanced concepts such as performing predictive analytics. You will also learn how to use Alteryx’s features to share the insights gained with the relevant decision makers. To ensure consistency, we will be using data from the Healthcare domain throughout this course. The knowledge you gain from this course will guide you to solve real-life problems related to Business Intelligence confidently. Whether you are a novice with Alteryx or an experienced data analyst keen to explore Alteryx’s self-service analytics features, this course will be the perfect companion for you.

Working in a hands-on learning environment, led by our Alteryx expert instructor, students will learn about and explore:

* Experience the power of codeless analytics using Alteryx, a leading Business Intelligence tool
* Uncover hidden trends and valuable insights from your data across different sources and make accurate predictions
* Includes real-world examples to put your understanding of the features in Alteryx to practical use.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Create efficient workflows with Alteryx to answer complex business questions
* Learn how to speed up the cleansing, data preparing, and shaping process
* Blend and join data into a single dataset for self-service analysis
* Write advanced expressions in Alteryx leading to an optimal workflow for efficient processing of huge data
* Develop high-quality, data-driven reports to improve consistency in reporting and analysis
* Explore the flexibility of macros by automating analytic processes
* Apply predictive analytics from spatial, demographic, and behavioral analysis and quickly publish, schedule
* Share your workflows and insights with relevant stakeholders

**Audience & Pre-Requisites**

This course is geared for attendees who wish to Implement a Business Intelligence solutions without any coding - by leveraging the power of the Alteryx platform

**Pre-Requisites:** Students should have

* developers with some knowledge of Alteryx and IT.

**Course Agenda / Topics**

1. **Getting Started with Alteryx**

* Getting Started with Alteryx
* Installation and setup
* The Alteryx Designer architecture
* Introduction to Alteryx Designer
* An overview of the workflow configurations
* Exploring the Tool Palettes
* The Favorites tools

1. **Workflow Optimization**

* Workflow Optimization
* Resource optimization and speed processing
* Performance Profiling
* Connecting to data
* Alteryx file types

1. **Data Preparation and Blending**

* Data Preparation and Blending
* Data preparation
* Data cleansing
* Filtering
* Join and union

1. **Writing Fast and Accurate Expressions**

* Writing Fast and Accurate Expressions
* Formula
* Multi-Row Formula
* Multi-Field Formula

1. **Transforming Data**

* Transforming Data
* Transforming data
* Summarizing and aggregating data
* Running total
* Weighted average

1. **Data Parsing Techniques**

* Data Parsing Techniques
* Text to columns
* Converting strings to dates and dates to strings
* Regular expressions

1. **Creating Data-Driven Custom Reports**

* Creating Data-Driven Custom Reports
* Data Table
* Charting
* Reporting layout
* Sharing the report

1. **Using Macros in Workflows**

* Using Macros in Workflows
* Standard Macro
* Batch Macro

1. **Sharing Your Insights**

* Sharing Your Insights
* Alteryx Server
* Alteryx Analytics Gallery

1. **Best Practices**

* Best Practices
* The ultimate guide to workflow design
* Tips and tricks

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Haskell Data Analysis

**Course Snapshot**

* **Course:** Haskell Data Analysis
* **Duration:** 2 days
* **Skill-level**: Foundation-level Haskell Data Analysis skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to get Haskell skills to work and generate publication-ready visualizations in no time at all.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Every business and organization that collects data is capable of tapping into its own data to gain insights how to improve. Haskell is a purely functional and lazy programming language, well-suited to handling large data analysis problems. This course will take you through the more difficult problems of data analysis in a hands-on manner. This course will help you get up-to-speed with the basics of data analysis and approaches in the Haskell language. You'll learn about statistical computing, file formats (CSV and SQLite3), descriptive statistics, charts, and progress to more advanced concepts such as understanding the importance of normal distribution. While mathematics is a big part of data analysis, we've tried to keep this course simple and approachable so that you can apply what you learn to the real world. By the end of this course, you will have a thorough understanding of data analysis, and the different ways of analyzing data. You will have a mastery of all the tools and techniques in Haskell for effective data analysis.

Working in a hands-on learning environment, led by our Haskell Data Analysis expert instructor, students will learn about and explore:

* Take your data analysis skills to the next level using the power of Haskell
* Understand regression analysis, perform multivariate regression, and untangle different cluster varieties
* Create publication-ready visualizations of data

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn to parse a CSV file and read data into the Haskell environment
* Create Haskell functions for common descriptive statistics functions
* Create an SQLite3 database using an existing CSV file
* Learn the versatility of SELECT queries for slicing data into smaller chunks
* Apply regular expressions in large-scale datasets using both CSV and SQLite3 files
* Create a Kernel Density Estimator visualization using normal distribution

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to get Haskell skills to work and generate publication-ready visualizations in no time at all.

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **Descriptive Statistics**

* Descriptive Statistics
* The CSV library – working with CSV files
* Data range
* Data mean and standard deviation
* Data median
* Data mode

1. **SQLite3**

* SQLite3
* SQLite3 command line
* Working with SQLite3 and Haskell
* Slices of data
* Working with SQLite3 and descriptive statistics

1. **Regular Expressions**

* Regular Expressions
* Dots and pipes
* Atom and Atom modifiers
* Character classes
* Regular expressions in CSV files
* SQLite3 and regular expressions

1. **Visualizations**

* Visualizations
* Line plots of a single variable
* Plotting a moving average
* Creating publication-ready plots
* Feature scaling
* Scatter plots

1. **Kernel Density Estimation**

* Kernel Density Estimation
* The central limit theorem
* Normal distribution
* Introducing kernel density estimation
* Application of the KDE

1. **Course Review**

* Course Review
* Converting CSV variation files into SQLite3
* Using SQLite3 SELECT and the DescriptiveStats module for descriptive statistics
* Creating compelling visualizations using EasyPlot
* Reintroducing kernel density estimation

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Python Feature Engineering Cookbook

**Course Snapshot**

* **Course:** Python Feature Engineering Cookbook
* **Duration:** 3 days
* **Skill-level**: Foundation-level Python Feature Engineering Cook skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Extract accurate information from data to train and improve machine learning models using NumPy, SciPy, pandas, and scikit-learn libraries.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Feature engineering is invaluable for developing and enriching your machine learning models. In this cookbook, you will work with the best tools to streamline your feature engineering pipelines and techniques and simplify and improve the quality of your code.

Using Python libraries such as pandas, scikit-learn, Feature tools, and Feature-engine, you’ll learn how to work with both continuous and discrete datasets and be able to transform features from unstructured datasets. You will develop the skills necessary to select the best features as well as the most suitable extraction techniques. This course will cover Python recipes that will help you automate feature engineering to simplify complex processes. You’ll also get to grips with different feature engineering strategies, such as the box-cox transform, power transform, and log transform across machine learning, reinforcement learning, and natural language processing (NLP) domains. By the end of this course, you’ll have discovered tips and practical solutions to all of your feature engineering problems.

Working in a hands-on learning environment, led by our Python expert instructor, students will learn about and explore:

* Discover solutions for feature generation, feature extraction, and feature selection
* Uncover the end-to-end feature engineering process across continuous, discrete, and unstructured datasets
* Implement modern feature extraction techniques using Python's pandas, scikit-learn, SciPy and NumPy libraries

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Simplify your feature engineering pipelines with powerful Python packages
* Get to grips with imputing missing values
* Encode categorical variables with a wide set of techniques
* Extract insights from text quickly and effortlessly
* Develop features from transactional data and time series data
* Derive new features by combining existing variables
* Understand how to transform, discretize, and scale your variables
* Create informative variables from date and time

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Extract accurate information from data to train and improve machine learning models using NumPy, SciPy, pandas, and scikit-learn libraries

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Foreseeing Variable Problems When Building ML Models**

* Foreseeing Variable Problems When Building ML Models
* Technical requirements
* Identifying numerical and categorical variables
* Quantifying missing data
* Determining cardinality in categorical variables
* Pinpointing rare categories in categorical variables
* Identifying a linear relationship
* Identifying a normal distribution
* Distinguishing variable distribution
* Highlighting outliers
* Comparing feature magnitude

1. **Imputing Missing Data**

* Imputing Missing Data
* Technical requirements
* Removing observations with missing data
* Performing mean or median imputation
* Implementing mode or frequent category imputation
* Replacing missing values with an arbitrary number
* Capturing missing values in a bespoke category
* Replacing missing values with a value at the end of the distribution
* Implementing random sample imputation
* Adding a missing value indicator variable
* Performing multivariate imputation by chained equations
* Assembling an imputation pipeline with scikit-learn
* Assembling an imputation pipeline with Feature-engine

1. **Encoding Categorical Variables**

* Encoding Categorical Variables
* Technical requirements
* Creating binary variables through one-hot encoding
* Performing one-hot encoding of frequent categories
* Replacing categories with ordinal numbers
* Replacing categories with counts or frequency of observations
* Encoding with integers in an ordered manner
* Encoding with the mean of the target
* Encoding with the Weight of Evidence
* Grouping rare or infrequent categories
* Performing binary encoding
* Performing feature hashing

1. **Transforming Numerical** **Variables**

* Transforming Numerical Variables
* Technical requirements
* Transforming variables with the logarithm
* Transforming variables with the reciprocal function
* Using square and cube root to transform variables
* Using power transformations on numerical variables
* Performing Box-Cox transformation on numerical variables
* Performing Yeo-Johnson transformation on numerical variables

1. **Performing Variable Discretization**

* Performing Variable Discretization
* Technical requirements
* Dividing the variable into intervals of equal width
* Sorting the variable values in intervals of equal frequency
* Performing discretization followed by categorical encoding
* Allocating the variable values in arbitrary intervals
* Performing discretization with k-means clustering
* Using decision trees for discretization

1. **Working with Outliers**

* Working with Outliers
* Technical requirements
* Trimming outliers from the dataset
* Performing winsorization
* Capping the variable at arbitrary maximum and minimum values
* Performing zero-coding – capping the variable at zero

1. **Deriving Features from Dates and Time Variables**

* Deriving Features from Dates and Time Variables
* Technical requirements
* Extracting date and time parts from a datetime variable
* Deriving representations of the year and month
* Creating representations of day and week
* Extracting time parts from a time variable
* Capturing the elapsed time between datetime variables
* Working with time in different time zones

1. **Performing Feature Scaling**

* Performing Feature Scaling
* Technical requirements
* Standardizing the features
* Performing mean normalization
* Scaling to the maximum and minimum values
* Implementing maximum absolute scaling
* Scaling with the median and quantiles
* Scaling to vector unit length

1. **Applying Mathematical Computations to Features**

* Applying Mathematical Computations to Features
* Technical requirements
* Combining multiple features with statistical operations
* Combining pairs of features with mathematical functions
* Performing polynomial expansion
* Deriving new features with decision trees
* Carrying out PCA

1. **Creating Features with Transactional and Time Series Data**

* Creating Features with Transactional and Time Series Data
* Technical requirements
* Aggregating transactions with mathematical operations
* Aggregating transactions in a time window
* Determining the number of local maxima and minima
* Deriving time elapsed between time-stamped events
* Creating features from transactions with Feature tools

1. **Extracting Features from Text Variables**

* Extracting Features from Text Variables
* Technical requirements
* Counting characters, words, and vocabulary
* Estimating text complexity by counting sentences
* Creating features with bag-of-words and n-grams
* Implementing term frequency-inverse document frequency
* Cleaning and stemming text variables

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# SciPy Recipes

**Course Snapshot**

* **Course:** SciPy Recipes
* **Duration:** 3 days
* **Skill-level**: Foundation-level SciPy Recipes skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Tackle the most sophisticated problems associated with scientific computing and data manipulation using SciPy.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

With the SciPy Stack, you get the power to effectively process, manipulate, and visualize your data using the popular Python language. Utilizing SciPy correctly can sometimes be a very tricky proposition. This course provides the right techniques so you can use SciPy to perform different data science tasks with ease. This course includes hands-on recipes for using the different components of the SciPy Stack such as NumPy, SciPy, matplotlib, and pandas, among others. You will use these libraries to solve real-world problems in linear algebra, numerical analysis, data visualization, and much more. The recipes included in the course will ensure you get a practical understanding not only of how a particular feature in SciPy Stack works, but also of its application to real-world problems. The independent nature of the recipes also ensure that you can pick up any one and learn about a particular feature of SciPy without reading through the other recipes, thus making the course a very handy and useful guide.

Working in a hands-on learning environment, led by our SciPy Recipes expert instructor, students will learn about and explore:

* Covers a wide range of data science tasks using SciPy, NumPy, pandas, and matplotlib
* Effective recipes on advanced scientific computations, statistics, data wrangling, data visualization, and more
* A must-have course if you're looking to solve your data-related problems using SciPy, on-the-go

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get a solid foundation in scientific computing using Python
* Master common tasks related to SciPy and associated libraries such as NumPy, pandas, and matplotlib
* Perform mathematical operations such as linear algebra and work with the statistical and probability functions in SciPy
* Master advanced computing such as Discrete Fourier Transform and K-means with the SciPy Stack
* Implement data wrangling tasks efficiently using pandas
* Visualize your data through various graphs and charts using matplotlib

**Audience & Pre-Requisites**

This course is geared for attendees who wish to Tackle the most sophisticated problems associated with scientific computing and data manipulation using SciPy

**Pre-Requisites:** Students should have

* developers with some knowledge of Python and IT skills.

**Course Agenda / Topics**

1. **Getting to Know the Tools**

* Getting to Know the Tools
* Introduction
* Installing Anaconda on Windows
* Installing Anaconda on macOS
* Installing Anaconda on Linux
* Checking the Anaconda installation
* Installing SciPy from a binary distribution on Windows
* Installing SciPy from a binary distribution on macOS
* Installing SciPy from source on Linux
* Installing optional packages with conda
* Installing packages with pip
* Setting up a virtual environment with conda
* Creating a virtual environment for development with conda
* Creating a conda environment with a different version of a package
* Using conda environments to run different versions of Python
* Creating virtual environments with venv
* Running SciPy in a script
* Running SciPy in Jupyter
* Running SciPy in Spyder
* Running SciPy in PyCharm

1. **Getting Started with NumPy**

* Getting Started with NumPy
* Introduction
* Creating NumPy arrays
* Querying and changing the shape of an array
* Storing and retrieving NumPy arrays
* Indexing
* Operations on arrays
* Using masked arrays to represent invalid data
* Using object arrays to store heterogeneous data
* Defining, symbolically, a function operating on arrays

1. **Using Matplotlib to Create Graphs**

* Using Matplotlib to Create Graphs
* Introduction
* Creating two-dimensional plots of functions and data
* Generating multiple plots in a single figure
* Setting line styles and markers
* Using different backends to display graphs
* Saving plots to disk
* Annotating graphs
* Generating histograms and box plots
* Creating three-dimensional plots
* Generating interactive displays in the Jupyter Notebook
* Object-oriented graph creation using Artist objects
* Creating a map with cartopy

1. **Data Wrangling with pandas**

* Data Wrangling with pandas
* Creating Series objects
* Creating DataFrame objects
* Inserting and deleting columns to a DataFrame
* Inserting and deleting rows to a DataFrame
* Selecting items by row indexes and column labels
* Selecting items by integer location
* Selecting items using mixed indexing
* Accessing, selecting, and modifying data
* Selecting rows using Boolean selection
* Reading and storing data in different formats
* Data displays employing different kinds of visual representation
* How to apply numerical functions and operations to Series and DataFrame objects
* Computing statistical functions on Series and DataFrame objects
* How to sort data in Series and DataFrame objects
* Performing merging, joins, concatenation, and grouping

1. **Matrices and Linear Algebra**

* Matrices and Linear Algebra
* Introduction
* Matrix operations and functions on two-dimensional arrays
* Solving linear systems using matrices
* Calculating the null space of a matrix
* Calculating the LU decompositions of a matrix
* Calculating the QR decomposition of a matrix
* Calculating the eigenvalue and eigenvector of a matrix
* Diagonalizing a matrix
* Calculating the Jordan form of a matrix
* Calculating the singular value decomposition of a matrix
* Creating a sparse matrix
* Computations on top of a sparse matrix

1. **Solving Equations and Optimization**

* Solving Equations and Optimization
* Introduction
* Non-linear equations and systems
* System of equations and how to solve it
* Choosing the solver used to find the solution of equations
* Solving constrained non-linear optimization problems in several variables
* Solving one-dimensional optimization problems
* Solving multidimensional non-linear equations using the Newton-Krylov method
* Solving multidimensional non-linear equations using the Anderson method
* Finding the best linear fit for a set of data
* Doing non-linear regression for a set of data
* Regression

1. **Constants and Special Functions**

* Constants and Special Functions
* Introduction
* Physical and mathematical constants available in SciPy
* Using constants in the CODATA database
* Bessel functions
* Error functions
* Orthogonal polynomials functions
* Gamma function
* The Riemann zeta function
* Airy and Bairy functions
* The Bessel and Struve functions

1. **Calculus, Interpolation, and Differential Equations**

* Calculus, Interpolation, and Differential Equations
* Introduction
* Integration
* Computing integrals using a Gaussian quadrature
* Computing integrals with weighting functions
* Computing multiple integrals
* Interpolation
* Computing a polynomial interpolation for a set of data points
* Univariate interpolation
* Finding a cubic spline that interpolates a set of data
* Defining a B-spline for a given set of control points
* Differentiation
* Solving a one-dimensional ordinary differential equation
* Solving a system of ordinary differential equations
* Solving differential equations and systems with parameters
* Using ode and the objected-oriented interface to solve differential equations

1. **Statistics and Probability**

* Statistics and Probability
* Introduction
* Computing the probability mass function of a discrete random variable
* Computing the probability density function of a continuous random variable
* Computing the cumulative distribution function for a random variable
* Computing the values of inverse probabilities associated with a random variable
* Computing the average, standard deviation, and higher moments of a distribution
* Computing probabilities associated with the multivariate Gaussian distribution
* Computing the summary statistics of a dataset

1. **Advanced Computations with SciPy**

* Advanced Computations with SciPy
* Discrete Fourier transforms
* Computing the discrete Fourier transform (DFT) of a data series using the FFT algorithm
* Computing the inverse DFT of a data series
* Computing signal construction
* Getting started with filters
* Computing the DFT for two-dimensional data
* How to find the DFT of the derivative of a function
* Computing the convolution of two functions
* Mathematical imaging
* Computing pairwise distances from a dataset, using different distance metrics
* How to identify neighborhoods and nearest neighbors for a dataset and a given metric
* Nearest neighbors regression

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Exploratory Data Analysis with Python

**Course Snapshot**

* **Course:** Exploratory Data Analysis with Python
* **Duration:** 4 days
* **Skill-level**: Foundation-level Exploratory Data Analysis with Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to use
* Discover techniques to summarize the characteristics of your data using PyPlot, NumPy, SciPy, and pandas.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Exploratory Data Analysis (EDA) is an approach to data analysis that involves the application of diverse techniques to gain insights into a dataset. This course will help you gain practical knowledge of the main pillars of EDA - data cleaning, data preparation, data exploration, and data visualization. You’ll start by performing EDA using open source datasets and perform simple to advanced analyses to turn data into meaningful insights. You’ll then learn various descriptive statistical techniques to describe the basic characteristics of data and progress to performing EDA on time-series data. As you advance, you’ll learn how to implement EDA techniques for model development and evaluation and build predictive models to visualize results. Using Python for data analysis, you’ll work with real-world datasets, understand data, summarize its characteristics, and visualize it for business intelligence. By the end of this EDA course, you’ll have developed the skills required to carry out a preliminary investigation on any dataset, yield insights into data, present your results with visual aids, and build a model that correctly predicts future outcomes

Working in a hands-on learning environment, led by our Data Analysis with Python expert instructor, students will learn about and explore:

* Understand the fundamental concepts of exploratory data analysis using Python
* Find missing values in your data and identify the correlation between different variables
* Practice graphical exploratory analysis techniques using Matplotlib and the Seaborn Python package

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Import, clean, and explore data to perform preliminary analysis using powerful Python packages
* Identify and transform erroneous data using different data wrangling techniques
* Explore the use of multiple regression to describe non-linear relationships
* Discover hypothesis testing and explore techniques of time-series analysis
* Understand and interpret results obtained from graphical analysis
* Build, train, and optimize predictive models to estimate results
* Perform complex EDA techniques on open source datasets

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Discover techniques to summarize the characteristics of your data using PyPlot, NumPy, SciPy, and pandas

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Section 1: The Fundamentals of EDA**

* Section 1: The Fundamentals of EDA

1. **Exploratory Data Analysis Fundamentals**

* Exploratory Data Analysis Fundamentals
* Understanding data science
* The significance of EDA
* Making sense of data
* Comparing EDA with classical and Bayesian analysis
* Software tools available for EDA
* Getting started with EDA

1. **Visual Aids for EDA**

* Visual Aids for EDA
* Technical requirements
* Line chart
* Bar charts
* Scatter plot
* Area plot and stacked plot
* Pie chart
* Table chart
* Polar chart
* Histogram
* Lollipop chart
* Choosing the best chart
* Other libraries to explore

1. **EDA with Personal Email**

* EDA with Personal Email
* Technical requirements
* Loading the dataset
* Data transformation
* Data analysis

1. **Data Transformation**

* Data Transformation
* Technical requirements
* Background
* Merging database-style dataframes
* Transformation techniques
* Benefits of data transformation

1. **Section 2: Descriptive Statistics**

* Section 2: Descriptive Statistics

1. **Descriptive Statistics**

* Descriptive Statistics
* Technical requirements
* Understanding statistics
* Measures of central tendency
* Measures of dispersion

1. **Grouping Datasets**

* Grouping Datasets
* Technical requirements
* Understanding groupby()
* Groupby mechanics
* Data aggregation
* Pivot tables and cross-tabulations

1. **Correlation**

* Correlation
* Technical requirements
* Introducing correlation
* Types of analysis
* Discussing multivariate analysis using the Titanic dataset
* Outlining Simpson's paradox
* Correlation does not imply causation

1. **Time Series Analysis**

* Time Series Analysis
* Technical requirements
* Understanding the time series dataset
* TSA with Open Power System Data

1. **Section 3: Model Development** **and Evaluation**

* Section 3: Model Development and Evaluation

1. **Hypothesis Testing and Regression**

* Hypothesis Testing and Regression
* Technical requirements
* Hypothesis testing
* p-hacking
* Understanding regression
* Model development and evaluation

1. **Model Development and Evaluation**

* Model Development and Evaluation
* Technical requirements
* Types of machine learning
* Understanding supervised learning
* Understanding unsupervised learning
* Understanding reinforcement learning
* Unified machine learning workflow

1. **EDA on Wine Quality Data Analysis**

* EDA on Wine Quality Data Analysis
* Technical requirements
* Disclosing the wine quality dataset
* Analyzing red wine
* Analyzing white wine
* Model development and evaluation

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# SAS for Data Analysis

**Course Snapshot**

* **Course:** SAS for Data Analysis
* **Duration:** 3 days
* **Skill-level**: Foundation-level SAS for Data Analysis skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to get Leverage the full potential of SAS to get unique, actionable insights from your data.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

SAS is one of the leading enterprise tools in the world today when it comes to data management and analysis. It enables the fast and easy processing of data and helps you gain valuable business insights for effective decision-making. This course will serve as a comprehensive guide that will prepare you for the SAS certification exam. After a quick overview of the SAS architecture and components, the course will take you through the different approaches to importing and reading data from different sources using SAS. You will then cover SAS Base and 4GL, understanding data management and analysis, along with exploring SAS functions for data manipulation and transformation. Next, you'll discover SQL procedures and get up to speed on creating and validating queries. In the concluding lessons, you'll learn all about data visualization, right from creating bar charts and sample geographic maps through to assigning patterns and formats. In addition to this, the course will focus on macro programming and its advanced aspects. By the end of this course, you will be well versed in SAS programming and have the skills you need to easily handle and manage your data-related problems in SAS.

Working in a hands-on learning environment, led by our SAS for Data Analysis expert instructor, students will learn about and explore:

* Build enterprise-class data solutions using SAS and become well-versed in SAS programming
* Work with different data structures, and run SQL queries to manipulate your data
* Explore essential concepts and techniques with practical examples to confidently pass the SAS certification exam

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Explore a variety of SAS modules and packages for efficient data analysis
* Use SAS 4GL functions to manipulate, merge, sort, and transform data
* Gain useful insights into advanced PROC SQL options in SAS to interact with data
* Get to grips with SAS Macro and define your own macros to share data
* Discover the different graphical libraries to shape and visualize data with
* Apply the SAS Output Delivery System to prepare detailed reports

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to get Leverage of full potential of SAS to get unique, actionable insights from your data

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Introduction to SAS Programming**

* Introduction to SAS Programming
* SAS dataset fundamentals
* SAS programming language – basic syntax
* SAS LOG
* Dataset options
* SAS operators
* Formats
* Subsetting datasets
* Dictionary tables

1. **Data Manipulation and Transformation**

* Data Manipulation and Transformation
* Length of a variable
* Case conversion and alignment
* String identification
* Dealing with blanks
* Missing and multiple values
* Interval calculations
* Concatenation
* Logic and control
* Number manipulation

1. **Combining, Indexing, Encryption, and Compression Techniques Simplified**

* Combining, Indexing, Encryption, and Compression Techniques Simplified
* Introduction to combining
* Concatenation
* Interleaving
* Merging
* Indexing
* Encryption

1. **Power of Statistics,** **Reporting, Transforming Procedures, and Functions**

* Power of Statistics, Reporting, Transforming Procedures, and Functions
* Proc Freq
* Proc Univariate
* Proc Means and Summary
* Proc Corr
* Proc REG
* Proc Transpose

1. **Advanced Programming Techniques - SAS Macros**

* Advanced Programming Techniques - SAS Macros
* What are macros?
* Macro variable processing
* Macro resolution tracking
* Macro definition processing
* Comparing positional and keywords parameters
* Data-driven programming
* Leveraging automatic global macro variables
* Macros that evaluate
* Writing efficient macros

1. **Powerful Functions, Options, and Automatic Variables Simplified**

* Powerful Functions, Options, and Automatic Variables Simplified
* NOMPREPLACE and MREPLACE
* NOMCOMPILE and NCOMPILE
* MCOMPILENOTE
* NOMEXECNOTE and MEXECNOTE
* MAUTOCOMPLOC
* MACRO and NOMACRO
* Exchanging values between the DATA step and macro variables
* CALL EXECUTE
* Altering the CALL SYMPUT example
* Resolving macro variables
* Macro quoting

1. **Advanced Programming Techniques Using PROC SQL**

* Advanced Programming Techniques Using PROC SQL
* Comparing data steps and Proc SQL
* Proc SQL joins
* Proc SQL essentials
* Dictionary tables

1. **Deep Dive into PROC SQL**

* Deep Dive into PROC SQL
* SAS views in Proc SQL
* Making changes with Proc SQL
* Identifying duplicates using Proc SQL
* Creating an index in Proc SQL
* Macros and Proc SQL

1. **Data Visualization**

* Data Visualization
* The role of data visualization in analytics
* Histograms
* Line plots
* Vertical and horizontal bar charts
* Scatter charts
* Box plot

1. **Reporting and Output Delivery System**

* Reporting and Output Delivery System
* Proc Tabulate
* Specifying the ODS destination
* Formatting ODS files
* ODS Excel charts

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# R Data Analysis Projects

**Course Snapshot**

* **Course:** R Data Analysis Projects
* **Duration:** 2 days
* **Skill-level**: Foundation-level R Data Analysis Projects skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Get valuable insights from your data by building data analysis systems from scratch with R..
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

R offers a large variety of packages and libraries for fast and accurate data analysis and visualization. As a result, it’s one of the most popularly used languages by data scientists and analysts, or anyone who wants to perform data analysis. This course will demonstrate how you can put to use your existing knowledge of data analysis in R to build highly efficient, end-to-end data analysis pipelines without any hassle. You’ll start by building a content-based recommendation system, followed by building a project on sentiment analysis with tweets. You’ll implement time-series modeling for anomaly detection, and understand cluster analysis of streaming data. You’ll work through projects on performing efficient market data research, building recommendation systems, and analyzing networks accurately, all provided with easy to follow codes. With the help of these real-world projects, you’ll get a better understanding of the challenges faced when building data analysis pipelines, and see how you can overcome them without compromising on the efficiency or accuracy of your systems. The course covers some popularly used R packages such as dplyr, ggplot2, RShiny, and others, and includes tips on using them effectively. By the end of this course, you’ll have a better understanding of data analysis with R, and be able to put your knowledge to practical use without any hassle.

Working in a hands-on learning environment, led by our R Data Analysis Projects expert instructor, students will learn about and explore:

* A handy guide to take your understanding of data analysis with R to the next level
* Real-world projects that focus on problems in finance, network analysis, social media, and more
* From data manipulation to analysis to visualization in R, this course will teach you everything you need to know about building end-to-end data analysis pipelines using R

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Build end-to-end predictive analytics systems in R
* Build an experimental design to gather your own data and conduct analysis
* Build a recommender system from scratch using different approaches
* Use and leverage RShiny to build reactive programming applications
* Build systems for varied domains including market research, network analysis, social media analysis, and more
* Explore various R Packages such as RShiny, ggplot, recommenderlab, dplyr, and find out how to use them effectively
* Communicate modeling results using Shiny Dashboards
* Perform multi-variate time-series analysis prediction, supplemented with sensitivity analysis and risk modeling

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Get valuable insights from your data by building data analysis systems from scratch with R.

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **Association Rule Mining**

* Association Rule Mining
* Understanding the recommender systems
* Retailer use case and data
* Association rule mining
* The cross-selling campaign
* Weighted association rule mining
* Hyperlink-induced topic search (HITS)
* Negative association rules
* Rules visualization
* Wrapping up

1. **Fuzzy Logic Induced Content-Based Recommendation**

* Fuzzy Logic Induced Content-Based Recommendation
* Introducing content-based recommendation
* News aggregator use case and data
* Designing the content-based recommendation engine
* Complete R Code

1. **Collaborative Filtering**

* Collaborative Filtering
* Collaborative filtering
* Recommenderlab package
* Use case and data
* Designing and implementing collaborative filtering
* Complete R Code

1. **Taming Time Series Data Using Deep Neural Networks**

* Taming Time Series Data Using Deep Neural Networks
* Time series data
* Deep neural networks
* Introduction to the MXNet R package
* Symbolic programming in MXNet
* Training test split
* Complete R code

1. **Twitter Text Sentiment Classification Using Kernel Density Estimates**

* Twitter Text Sentiment Classification Using Kernel Density Estimates
* Kernel density estimation
* Twitter text
* Sentiment classification
* Dictionary based scoring
* Text pre-processing
* Building a sentiment classifier
* Assembling an RShiny application
* Complete R code

1. **Record Linkage - Stochastic and Machine Learning Approaches**

* Record Linkage - Stochastic and Machine Learning Approaches
* Introducing our use case
* Demonstrating the use of RecordLinkage package
* Stochastic record linkage
* Machine learning-based record linkage
* Building an RShiny application
* Complete R code

1. **Streaming Data Clustering Analysis in R**

* Streaming Data Clustering Analysis in R
* Streaming data and its challenges
* Introducing stream clustering
* Introducing the stream package
* Use case and data
* Complete R code

1. **Analyze and Understand Networks Using R**

* Analyze and Understand Networks Using R
* Graphs in R
* Use case and data
* Data preparation
* Product network analysis
* Building a RShiny application
* The complete R script

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Analysis with Pandas

**Course Snapshot**

* **Course:** Data Analysis with Pandas
* **Duration:** 3 days
* **Skill-level**: Foundation-level Data Analysis with Pandas skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Get grips with pandas - a versatile and high-performance Python library for data manipulation, analysis, and discovery.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Data analysis has become a necessary skill in a variety of domains where knowing how to work with data and extract insights can generate significant value. Data Analysis with Pandas will show you how to analyze your data, get started with machine learning, and work effectively with Python libraries often used for data science, such as pandas, NumPy, matplotlib, seaborn, and scikit-learn. Using real-world datasets, you will learn how to use the powerful pandas library to perform data wrangling to reshape, clean, and aggregate your data. Then, you will be able to conduct exploratory data analysis by calculating summary statistics and visualizing the data to find patterns. In the concluding lessons, you will explore some applications of anomaly detection, regression, clustering, and classification using scikit-learn to make predictions based on past data. By the end of this course, you will be equipped with the skills you need to use pandas to ensure the veracity of your data, visualize it for effective decision-making, and reliably reproduce analyses across multiple datasets.

Working in a hands-on learning environment, led by our SAS for Data Analysis expert instructor, students will learn about and explore:

* Perform efficient data analysis and manipulation tasks using pandas
* Apply pandas to different real-world domains with the help of step-by-step demonstrations
* Get accustomed to using pandas as an effective data exploration tool.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand how data analysts and scientists gather and analyze data
* Perform data analysis and data wrangling using Python
* Combine, group, and aggregate data from multiple sources
* Create data visualizations with pandas, matplotlib, and seaborn
* Apply machine learning (ML) algorithms to identify patterns and make predictions
* Use Python data science libraries to analyze real-world datasets
* Use pandas to solve common data representation and analysis problems
* Build Python scripts, modules, and packages for reusable analysis code

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to get Leverage of full potential of SAS to get unique, actionable insights from your data

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Introduction to Data Analysis**

* Introduction to Data Analysis
* Fundamentals of data analysis
* Statistical foundations
* Setting up a virtual environment

1. **Working with Pandas DataFrames**

* Working with Pandas DataFrames
* Pandas data structures
* Bringing data into a pandas DataFrame
* Inspecting a DataFrame object
* Grabbing subsets of the data
* Adding and removing data

1. **Data Wrangling with Pandas**

* Data Wrangling with Pandas
* What is data wrangling?
* Collecting temperature data
* Cleaning up the data
* Restructuring the data
* Handling duplicate, missing, or invalid data

1. **Aggregating Pandas DataFrames**

* Aggregating Pandas DataFrames
* Database-style operations on DataFrames
* DataFrame operations
* Aggregations with pandas and numpy
* Time series

1. **Visualizing Data with Pandas and Matplotlib**

* Visualizing Data with Pandas and Matplotlib
* An introduction to matplotlib
* Plotting with pandas
* The pandas.plotting subpackage

1. **Plotting with Seaborn and Customization Techniques**

* Plotting with Seaborn and Customization Techniques
* Utilizing seaborn for advanced plotting
* Formatting
* Customizing visualizations

1. **Financial Analysis - Bitcoin and the Stock Market**

* Financial Analysis - Bitcoin and the Stock Market
* Building a Python package
* Data extraction with pandas
* Exploratory data analysis
* Technical analysis of financial instruments
* Modeling performance

1. **Rule-Based Anomaly Detection**

* Rule-Based Anomaly Detection
* Simulating login attempts
* Exploratory data analysis
* Rule-based anomaly detection

1. **Getting Started with Machine Learning in Python**

* Getting Started with Machine Learning in Python
* Learning the lingo
* Exploratory data analysis
* Preprocessing data
* Clustering
* Regression
* Classification

1. **Making Better Predictions - Optimizing Models**

* Making Better Predictions - Optimizing Models
* Hyperparameter tuning with grid search
* Feature engineering
* Ensemble methods
* Inspecting classification prediction confidence
* Addressing class imbalance
* Regularization

1. **Machine Learning Anomaly Detection**

* Machine Learning Anomaly Detection
* Exploring the data
* Unsupervised methods
* Supervised methods
* Online learning

1. **The Road Ahead**

* The Road Ahead
* Data resources
* Practicing working with data
* Python practice

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Exploratory Data Analysis with R

**Course Snapshot**

* **Course:** Exploratory Data Analysis with R
* **Duration:** 3 days
* **Skill-level**: Foundation-level Exploratory Data Analysis with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Learn exploratory data analysis concepts using powerful R packages to enhance your R data analysis skills.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Exploratory Data Analysis with R will help you build not just a foundation but also expertise in the elementary ways to analyze data. You will learn how to understand your data and summarize its main characteristics. You'll also uncover the structure of your data, and you'll learn graphical and numerical techniques using the R language. This course covers the entire exploratory data analysis (EDA) process—data collection, generating statistics, distribution, and invalidating the hypothesis. As you progress through the course, you will learn how to set up a data analysis environment with tools such as ggplot2, knitr, and R Markdown, using tools such as DOE Scatter Plot and SML2010 for multifactor, optimization, and regression data problems. By the end of this course, you will be able to successfully carry out a preliminary investigation on any dataset, identify hidden insights, and present your results in a business context.

Working in a hands-on learning environment, led by our Data Analysis with R expert instructor, students will learn about and explore:

* Speed up your data analysis projects using powerful R packages and techniques
* Create multiple hands-on data analysis projects using real-world data
* Discover and practice graphical exploratory analysis techniques across domains

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn powerful R techniques to speed up your data analysis projects
* Import, clean, and explore data using powerful R packages
* Practice graphical exploratory analysis techniques
* Create informative data analysis reports using ggplot2
* Identify and clean missing and erroneous data
* Explore data analysis techniques to analyze multi-factor datasets

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Learn exploratory data analysis concepts using powerful R packages to enhance your R data analysis skills

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.
* experienced with spreadsheet software who know the basics of Python.

**Course Agenda / Topics**

1. **Setting Up Our Data Analysis Environment**

* Setting Up Our Data Analysis Environment
* Technical requirements
* The benefits of EDA across vertical markets
* Manipulating data
* Installing the required R packages and tools

1. **Importing Diverse Datasets**

* Importing Diverse Datasets
* Technical requirements
* Converting rectangular data into R with the readr R package
* Reading in Excel data with the readxl R package
* Reading in JSON data with the jsonlite R package
* Getting data into R from web APIs using the httr R package
* Getting data into R by scraping the web using the rvest package
* Importing data into R from relational databases using the DBI R package

1. **Examining, Cleaning, and Filtering**

* Examining, Cleaning, and Filtering
* Technical requirements
* About the dataset
* Reshaping and tidying up erroneous data
* Manipulating and mutating data
* Selecting and filtering data
* Cleaning and manipulating time series data

1. **Visualizing Data Graphically with ggplot2**

* Visualizing Data Graphically with ggplot2
* Technical requirements
* Advanced graphics grammar of ggplot2
* Installing ggplot2
* Scatter plots
* Histogram plots
* Density plots
* Probability plots
* Box plots
* Residual plots

1. **Creating Aesthetically Pleasing Reports with knitr and R Markdown**

* Creating Aesthetically Pleasing Reports with knitr and R Markdown
* Technical requirements
* Installing R Markdown
* Reproducible data analysis reports with knitr
* Exporting and customizing reports

1. **Univariate and Control Datasets**

* Univariate and Control Datasets
* Technical requirements
* Reading the dataset
* Cleaning and tidying up the data
* Understanding the structure of the data
* Hypothesis tests
* Tietjen-Moore test
* Parsimonious models
* Probability plots
* The Shapiro-Wilk test

1. **Time Series Datasets**

* Time Series Datasets
* Technical requirements
* Introducing and reading the dataset
* Cleaning the dataset
* Mapping and understanding structure
* Hypothesis test
* Grubbs' test and checking outliers
* Parsimonious models
* Bartlett's test
* Data visualization

1. **Multivariate Datasets**

* Multivariate Datasets
* Technical requirements
* Introducing and reading a dataset
* Cleaning the data
* Mapping and understanding the structure
* Hypothesis test
* Parsimonious model
* Levene's test
* Data visualization

1. **Multi-Factor Datasets**

* Multi-Factor Datasets
* Technical requirements
* Introducing and reading the dataset
* Cleaning the dataset
* Mapping and understanding data structure
* Hypothesis test
* Grubbs test and checking outliers
* Parsimonious model
* Multi-factor variance analysis
* Exploring graphically the dataset

1. **Handling Optimization and Regression Data Problems**

* Handling Optimization and Regression Data Problems
* Technical requirements
* Introducing and reading a dataset
* Cleaning the dataset
* Mapping and understanding the data structure
* Hypothesis test
* Grubbs' test and checking outliers
* Parsimonious model
* Exploration using graphics

1. **Next Steps**

* Next Steps
* Technical requirements
* What to learn next
* Why R?
* The data analysis workflow
* Building a data science portfolio
* Datasets in R
* Getting help with exploratory data analysis

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Data Analysis with R

**Course Snapshot**

* **Course:** Data Analysis with R
* **Duration:** 4 days
* **Skill-level**: Foundation-level Data Analysis with R skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for Python experienced developers, analysts or others who wants to Learn, by example, the fundamentals of data analysis as well as several intermediate to advanced methods and techniques ranging from classification and regression to Bayesian methods and MCMC, which can be put to immediate use.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Frequently the tool of choice for academics, R has spread deep into the private sector and can be found in the production pipelines at some of the most advanced and successful enterprises. The power and domain-specificity of R allows the user to express complex analytics easily, quickly, and succinctly. Starting with the basics of R and statistical reasoning, this course dives into advanced predictive analytics, showing how to apply those techniques to real-world data though with real-world examples. Packed with engaging problems and exercises, this course begins with a review of R and its syntax with packages like Rcpp, ggplot2, and dplyr. From there, get to grips with the fundamentals of applied statistics and build on this knowledge to perform sophisticated and powerful analytics. Solve the difficulties relating to performing data analysis in practice and find solutions to working with messy data, large data, communicating results, and facilitating reproducibility. This course is engineered to be an invaluable resource through many stages of anyone’s career as a data analyst.

Working in a hands-on learning environment, led by our Data Analysis with R expert instructor, students will learn about and explore:

* Analyze your data using R – the most powerful statistical programming language
* Learn how to implement applied statistics using practical use-cases
* Use popular R packages to work with unstructured and structured data

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Gain a thorough understanding of statistical reasoning and sampling theory
* Employ hypothesis testing to draw inferences from your data
* Learn Bayesian methods for estimating parameters
* Train regression, classification, and time series models
* Handle missing data gracefully using multiple imputation
* Identify and manage problematic data points
* Learn how to scale your analyses to larger data with Rcpp, data.table, dplyr, and parallelization
* Put best practices into effect to make your job easier and facilitate reproducibility

**Audience & Pre-Requisites**

This course is geared for attendees with Python skills who wish to Learn, by example, the fundamentals of data analysis as well as several intermediate to advanced methods and techniques ranging from classification and regression to Bayesian methods and MCMC, which can be put to immediate use.

**Pre-Requisites:** Students should have

* developers with some knowledge of Python.

**Course Agenda / Topics**

1. **RefresheR**

* RefresheR
* Navigating the basics
* Getting help in R
* Vectors
* Functions
* Matrices
* Loading data into R
* Working with packages

1. **The Shape of Data**

* The Shape of Data
* Univariate data
* Frequency distributions
* Central tendency
* Spread
* Populations, samples, and estimation
* Probability distributions
* Visualization methods

1. **Describing Relationships**

* Describing Relationships
* Multivariate data
* Relationships between a categorical and continuous variable
* Relationships between two categorical variables
* The relationship between two continuous variables
* Visualization methods

1. **Probability**

* Probability
* Basic probability
* A tale of two interpretations
* Sampling from distributions
* The normal distribution

1. **Using Data To Reason About The World**

* Using Data To Reason About The World
* Estimating means
* The sampling distribution
* Interval estimation
* Smaller samples

1. **Testing Hypotheses**

* Testing Hypotheses
* The null hypothesis significance testing framework
* Testing the mean of one sample
* Testing two means
* Testing more than two means
* Testing independence of proportions
* What if my assumptions are unfounded?

1. **Bayesian Methods**

* Bayesian Methods
* The big idea behind Bayesian analysis
* Choosing a prior
* Who cares about coin flips
* Enter MCMC – stage left
* Using JAGS and runjags
* Fitting distributions the Bayesian way
* The Bayesian independent samples t-test

1. **The Bootstrap**

* The Bootstrap
* What's... uhhh... the deal with the bootstrap?
* Performing the bootstrap in R (more elegantly)
* Confidence intervals
* A one-sample test of means
* Bootstrapping statistics other than the mean
* Busting bootstrap myths

1. **Predicting Continuous Variables**

* Predicting Continuous Variables
* Linear models
* Simple linear regression
* Simple linear regression with a binary predictor
* Multiple regression
* Regression with a non-binary predictor
* Kitchen sink regression
* The bias-variance trade-off
* Linear regression diagnostics
* Advanced topics

1. **Predicting Categorical Variables**

* Predicting Categorical Variables
* k-Nearest neighbors
* Logistic regression
* Decision trees
* Random forests
* Choosing a classifier

1. **Predicting Changes with Time**

* Predicting Changes with Time
* What is a time series?
* What is forecasting?
* Creating and plotting time series
* Components of time series
* Time series decomposition
* White noise
* Autocorrelation
* Smoothing
* ETS and the state space model
* Interventions for improvement
* What we didn't cover
* Citations for the climate change data

1. **Sources of Data**

* Sources of Data
* Relational databases
* Using JSON
* XML
* Other data formats
* Online repositories

1. **Dealing with Missing Data**

* Dealing with Missing Data
* Analysis with missing data
* Visualizing missing data
* Types of missing data
* Unsophisticated methods for dealing with missing data
* So how does mice come up with the imputed values?

1. **Dealing with Messy Data**

* Dealing with Messy Data
* Checking unsanitized data
* Regular expressions
* Other tools for messy data

1. **Dealing with Large Data**

* Dealing with Large Data
* Wait to optimize
* Using a bigger and faster machine
* Be smart about your code
* Using optimized packages
* Using another R implementation
* Using parallelization
* Using Rcpp
* Being smarter about your code

1. **Working with Popular R Packages**

* Working with Popular R Packages
* The data.table package
* Using dplyr and tidyr to manipulate data
* Functional programming as a main tidyverse principle
* Reshaping data with tidyr

1. **Reproducibility and Best Practices**

* Reproducibility and Best Practices
* R scripting
* R projects
* Version control
* Communicating results

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Exploring the Data Jungle

Course Snapshot

* **Course:** Exploring the Data Jungle
* **Duration:** 1 days
* **Skill-level**: Foundation level Data Jungle skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants create insight from chaos and discover important analytic patterns - to set your business on the right track..
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Exploring the Data Jungle:** Finding, Preparing, and Using Real-World Data is a collection of three hand-picked lessons introducing you to the often-overlooked art of putting unfamiliar data to good use. Brian Godsey, author of Think Like a Data Scientist, has selected these lessons to help you navigate data in the wild, identify and prepare raw data for analysis, modeling, machine learning, or visualization. As you explore the data jungle, you'll discover real-world examples in Python, R, and other languages suitable for data science.

Working in a hands-on learning environment, led by our Data Jungle expert instructor, students will learn about and explore:

* you need the right perspective and guidance. (There's no point hacking at overgrowth with a spoon after all!) Identify and prepare your data well
* you'll be well set to create insight from chaos and discover important analytic patterns - to set your business on the right track.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Data All Around Us: The Virtual Wilderness
* Exploring Data
* Real-world Data

Audience & Pre-Requisites

**Pre-Requisites:** Students should have

* Basic knowledge regarding it
* Familiar with programming skills

Course Agenda / Topics

1. **Data All Around Us: The Virtual Wildernessfree**

* Data as the object of study
* Where data might live, and how to interact with it
* Scouting for data
* Example: microRNA and gene expression

1. **Exploring Data**

* Using summary statistics to spot problems
* Spotting problems using graphics and visualization

1. **Real-world Data**

* Getting started: data collection
* Preprocessing the data for modeling
* Using data visualization
* Terms from this lesson

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Visualizing Graph Data

**Course Snapshot**

* **Course:** Visualizing Graph Data
* **Duration:** 3 days
* **Skill-level**: Foundation-level Visualizing Graph Data for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to turn their rich data into important insight? Complex relationships in large data sets can be difficult to recognize. Visualizing these connections as graphs makes it possible to see the patterns, so you can find meaning in an otherwise over-whelming sea of facts.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Visualizing Graph Data teaches you how to understand graph data, build graph data structures, and create meaningful visualizations. This engaging course gently introduces graph data visualization through fascinating examples and compelling case studies. You'll discover simple, but effective, techniques to model your data, handle big data, and depict temporal and spatial data. By the end, you'll have a conceptual foundation as well as the practical skills to explore your own data with confidence

Working in a hands-on learning environment, led by our Visualizing Graph Data instructor, students will learn about and explore:

* Visualizing Graph Data teaches you not only how to build graph data structures, but also how to create your own dynamic and interactive visualizations using a variety of tools.
* You'll discover simple, but effective, techniques to model your data, handle big data, and depict temporal and spatial data
* This course is loaded with fascinating examples and case studies to show you the real-world value of graph visualizations.
* You'll have a conceptual foundation as well as the practical skills to explore your own data with confidence

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* How to understand graph data structures, and create meaningful visualizations
* Techniques for creating effective visualizations
* Examples using the Gephi and Key Lines visualization packages
* Real-world case studies

**Audience & Pre-Requisites**

This course is geared for those who wants a conceptual foundation as well as the practical skills to explore your own data with confidence.

**Pre-Requisites:** Students should have

* No prior experience with graph data is required.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

**Part 1. Graph visualization basics**

1. **Getting to know graph visualization free**

* Getting to know graphs
* Getting to know graph visualization

1. **Case studies**

* Intelligence and terrorism
* Credit card fraud
* Cyber security
* Sales and marketing graphs

1. **An introduction to Gephi and Key Lines**

* Gephi
* Key Lines

**Part 2. Visualize your own data**

1. **Data modeling**

* What is a data model?
* Graph data models
* Graph databases

1. **How to build graph visualizations**

* Understanding your user
* Using intuitive visual properties
* Building charts with visual properties

1. **Creating interactive visualizations**

* Chart navigation
* Declutter your charts
* Data volumes
* Animations and mobile

1. **How to organize a chart**

* Force-directed layouts
* Other layout options

1. **Big data: using graphs when there is too much data**

* Controlling which nodes and edges are visible
* Grouping and combinations

1. **Dynamic graphs: how to show data over time free**

* How do graphs change over time?
* How to visualize changes over time
* Implementing dynamic graphs

1. **Graphs on maps: the where of graph visualization**

* Working with geographical data
* Overlaying graphs on maps
* Building graphs on maps

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Taming Text

**Course Snapshot**

* **Course:** Taming Text
* **Duration:** 2 days
* **Skill-level**: Foundation-level Taming Text skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to explores how to automatically organize text using approaches such as full-text search, proper name recognition, clustering, tagging, information extraction, and summarization
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

There is so much text in our lives, we are practically drowning in it. Fortunately, there are innovative tools and techniques for managing unstructured information that can throw the smart developer a much-needed lifeline. You'll find them in this course.

Taming Text is a practical, example-driven guide to working with text in real applications. This course introduces you to useful techniques like full-text search, proper name recognition, clustering, tagging, information extraction, and summarization. You'll explore real use cases as you systematically absorb the foundations upon which they are built.

Written in a clear and concise style, this course avoids jargon, explaining the subject in terms you can understand without a background in statistics or natural language processing. Examples are in Java, but the concepts can be applied in any language.

Working in a hands-on learning environment, led by our Taming Text instructor, students will learn about and explore:

* Taming Text is a hands-on, example-driven guide to working with unstructured text in the context of real-world applications.
* explores how to automatically organize text using approaches such as full-text search, proper name recognition, clustering, tagging, information extraction, and summarization.
* guides you through examples illustrating each of these topics, as well as the foundations upon which they are built.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* When to use text-taming techniques
* Important open-source libraries like Solr and Mahout
* How to build text-processing applications

**Audience & Pre-Requisites**

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, Microsoft azure and Machine Learning knowledge
* Good foundational mathematics or logic skills

**Course Agenda / Topics**

1. **Getting started taming textfree**

* Why taming text is important
* Preview: A fact-based question answering system
* Understanding text is hard
* Text, tamed
* Text and the intelligent app: search and beyond

1. **Foundations of taming text**

* Foundations of language
* Common tools for text processing
* Preprocessing and extracting content from common file formats

1. **Searching**

* Search and faceting example: Amazon.com
* Introduction to search concepts
* Introducing the Apache Solr search server
* Indexing content with Apache Solr
* Searching content with Apache Solr
* Understanding search performance factors
* Improving search performance
* Search alternatives

1. **Fuzzy string matching**

* Approaches to fuzzy string matching
* Finding fuzzy string matches
* Building fuzzy string matching applications

1. **Identifying people, places, and things**

* Approaches to named-entity recognition
* Basic entity identification with OpenNLP
* In-depth entity identification with OpenNLP
* Performance of OpenNLP
* Customizing OpenNLP entity identification for a new domain

1. **Clustering text**

* Google News document clustering
* Clustering foundations
* Setting up a simple clustering application
* Clustering search results using Carrot2
* Clustering document collections with Apache Mahout
* Topic modeling using Apache Mahout
* Examining clustering performance

1. **Classification, categorization, and tagging**

* Introduction to classification and categorization
* The classification process
* Building document categorizers using Apache Lucene
* Training a naive Bayes classifier using Apache Mahout
* Categorizing documents with OpenNLP
* Building a tag recommender using Apache Solr

1. **Building an example question answering systemfree**

* Basics of a question answering system
* Installing and running the QA code
* A sample question answering architecture
* Understanding questions and producing answers
* Steps to improve the system

1. **Untamed text: exploring the next frontier**

* Semantics, discourse, and pragmatics: exploring higher levels of NLP
* Document and collection summarization
* Relationship extraction
* Identifying important content and people
* Detecting emotions via sentiment analysis
* Cross-language information retrieval

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# D3.js

Course Snapshot

* **Course:** D3.js
* **Duration:** 3 days
* **Skill-level**: Foundation level D3.js skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to create interactive graphics and data-driven applications using D3.js. You'll start with in-depth explanations of D3's out-of-the-box layouts, along with dozens of practical use cases that align with different types of visualizations
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**D3.js** introduces you to the most powerful web data visualization library available and shows you how to use it to build interactive graphics and data-driven applications. You'll start with dozens of practical use cases that align with different types of charts, networks, and maps using D3's out-of-the-box layouts. Then, you'll explore practical techniques for content design, animation, and representation of dynamic data—including interactive graphics and live streaming data.

Working in a hands-on learning environment, led by our D3.js expert instructor, students will learn about and explore:

* you'll explore practical techniques for content creation, animation, and representing dynamic data—including interactive graphics and data streamed live over the web.
* The final lessons show you how to use D3's rich interaction model as the foundation for a complete web application. In the end
* you'll be ready to integrate D3.js into your web development process and transform any site into a more engaging and sophisticated user experience.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Interacting with vector graphics
* Expressive data visualization
* Creating rich mapping applications
* Prepping your data
* Complete data-driven web apps in D3

Audience & Pre-Requisites

This course is written for Readers have basic HTML, CSS, and JavaScript skills.

**Pre-Requisites:** Students should have

No experience with D3 or SVG is required

Course Agenda / Topics

1. **An introduction to D3.jsfree audio**

* What is D3.js?
* How D3 works
* Using HTML5
* Data standards
* Infoviz standards expressed in D3
* Your first D3 app

1. **Information visualization data flow**

* Working with data
* Data-binding
* Data presentation style, attributes, and content

1. **Data-driven design and interaction**

* Project architecture
* Interactive style and DOM
* Pregenerated content

1. **Chart components**

* General charting principles
* Creating an axis
* Complex graphical objects
* Line charts and interpolations
* Complex accessor functions

1. **Layouts free audio**

* Histograms
* Pie charts
* Pack layouts
* Trees
* Stack layout
* Plugins to add new layouts

1. **Network visualization**

* Static network diagrams
* Force-directed layout

1. **Geospatial information visualization**

* Basic mapmaking
* Better mapping
* Advanced mapping
* TopoJSON data and functionality
* Tile mapping with d3.geo.tile
* Further reading for web mapping

1. **Traditional DOM manipulation with D3**

* Setup
* Spreadsheet
* Canvas
* Image gallery

1. **Composing interactive applications**

* One data source, many perspectives
* Interactivity: hover events
* Brushing

1. **Writing layouts and components free audio**

* Creating a layout
* Writing your own components

1. **Big data visualization**

* Big geodata
* Big network data
* Optimizing xy data selection with quadtrees
* More optimization techniques

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Web Scraping with Python

Course Snapshot

* **Course:** Web Scraping with Python
* **Duration:** 3 days
* **Skill-level**: Foundation level Web Scraping with Python skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Collect and scrape different complexities of data from the modern Web using the latest tools, best practices, and techniques
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Web scraping is an essential technique used in many organizations to gather valuable data from web pages. This course will enable you to delve into web scraping techniques and methodologies. The course will introduce you to the fundamental concepts of web scraping techniques and how they can be applied to multiple sets of web pages. You'll use powerful libraries from the Python ecosystem such as Scrapy, lxml, pyquery, and bs4 to carry out web scraping operations. You will then get up to speed with simple to intermediate scraping operations such as identifying information from web pages and using patterns or attributes to retrieve information. This course adopts a practical approach to web scraping concepts and tools, guiding you through a series of use cases and showing you how to use the best tools and techniques to efficiently scrape web pages. You'll even cover the use of other popular web scraping tools, such as Selenium, Regex, and web-based APIs. By the end of this course, you will have learned how to efficiently scrape the web using different techniques with Python and other popular tools.

Working in a hands-on learning environment, led by our Web Scraping with Python expert instructor, students will learn about and explore:

* Learn different scraping techniques using a range of Python libraries such as Scrapy and Beautiful Soup
* Build scrapers and crawlers to extract relevant information from the web
* Automate web scraping operations to bridge the accuracy gap and manage complex business needs

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Analyze data and information from web pages
* Learn how to use browser-based developer tools from the scraping perspective
* Use XPath and CSS selectors to identify and explore markup elements
* Learn to handle and manage cookies
* Explore advanced concepts in handling HTML forms and processing logins
* Optimize web securities, data storage, and API use to scrape data
* Use Regex with Python to extract data
* Deal with complex web entities by using Selenium to find and extract data

Audience & Pre-Requisites

This course is written for Readers wants to Collect and scrape different complexities of data from the modern Web using the latest tools, best practices, and techniques.

**Pre-Requisites:** Students should have

Basic IT and Python knowledge and skills

Course Agenda / Topics

1. **Web Scraping Fundamentals**

* Web Scraping Fundamentals
* Introduction to web scraping
* Understanding web development and technologies
* Data finding techniques for the web
* Python and the Web – Using urllib and Requests
* Python and the Web – Using urllib and Requests
* Technical requirements
* Accessing the web with Python
* URL handling and operations with urllib and requests
* Implementing HTTP methods

1. **Using LXML, XPath, and CSS Selectors**

* Using LXML, XPath, and CSS Selectors
* Technical requirements
* Introduction to XPath and CSS selector
* Using web browser developer tools for accessing web content
* Scraping using lxml, a Python library

1. **Scraping Using pyquery – a Python Library**

* Scraping Using pyquery – a Python Library
* Technical requirements
* Introduction to pyquery
* Exploring pyquery
* Web scraping using pyquery

1. **Web Scraping Using Scrapy and Beautiful Soup**

* Web Scraping Using Scrapy and Beautiful Soup
* Technical requirements
* Web scraping using Beautiful Soup
* Web scraping using Scrapy
* Deploying a web crawler

1. **Working with Secure Web**

* Working with Secure Web
* Technical requirements
* Introduction to secure web
* HTML <form> processing
* Handling user authentication
* Working with cookies and sessions

1. **Data Extraction Using Web-Based APIs**

* Data Extraction Using Web-Based APIs
* Technical requirements
* Introduction to web APIs
* Accessing web API and data formats
* Web scraping using APIs

1. **Using Selenium to Scrape the Web**

* Using Selenium to Scrape the Web
* Technical requirements
* Introduction to Selenium
* Using Selenium for web scraping

1. **Using Regex to Extract Data**

* Using Regex to Extract Data
* Technical requirements
* Overview of regular expressions
* Using regular expressions to extract data

1. **Next Steps**

* Next Steps
* Technical requirements
* Managing scraped data
* Analysis and visualization using pandas and matplotlib
* Machine learning
* Data mining

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.



CATEGORY Big Data

# Redis.

Course Snapshot

* **Course:** Redis
* **Duration:** 3 days
* **Skill-level**: Foundation-level Redis skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to dive into real use cases including simple caching, distributed ad targeting, and more
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Redis** introduces Redis and walks you through examples that demonstrate how to use it effectively. You'll begin by getting Redis set up properly and then exploring the key-value model. Then, you'll dive into real use cases including simple caching, distributed ad targeting, and more. You'll learn how to scale Redis from small jobs to massive datasets. Experienced developers will appreciate lessons on clustering and internal scripting to make Redis easier to use.

Working in a hands-on learning environment, led by our Redis expert instructor, students will learn about and explore:

* Redis expands on the key-value pattern by accepting a wide variety of data types, including hashes, strings, lists, and other structures.
* It provides lightning-fast operations on in-memory datasets, and also makes it easy to persist to disk on the fly. Plus, it's free and open source.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Redis from the ground up
* Preprocessing real-time data
* Managing in-memory datasets
* Pub/sub and configuration
* Persisting to disk

Audience & Pre-Requisites

This course is geared for attendees who want to learn how to scale Redis from small jobs to massive datasets and Written for developers familiar with database concepts

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills
* No prior exposure to Redis or other NoSQL databases required.
* Appropriate for systems administrators comfortable with programming.
* Good foundational mathematics or logic skills

Course Agenda / Topics

1. **Getting to know Redisfree**

* What is Redis?
* What Redis data structures look like
* Hello Redis
* Getting help

1. **Anatomy of a Redis web application**

* Login and cookie caching
* Shopping carts in Redis
* Web page caching
* Database row caching
* Web page analytics

1. **Commands in Redis**

* Strings
* Lists
* Sets
* Hashes
* Sorted sets
* Publish/subscribe
* Other commands

1. **Keeping data safe and** **ensuring performancefree**

* Persistence options
* Replication
* Handling system failures
* Redis transactions
* Non-transactional pipelines
* Performance considerations

1. **Using Redis for application support**

* Logging to Redis
* Counters and statistics
* IP-to-city and -country lookup

1. **Service discovery and configuration**

* Application components in Redis
* Autocomplete
* Distributed locking
* Counting semaphores
* Task queues
* Pull messaging
* Distributing files with Redis

1. **Search-based applications**

* Searching in Redis
* Sorted indexes
* Ad targeting
* Job search

1. **Building a simple social networkfree**

* Users and statuses
* Home timeline
* Followers/following lists
* Posting or deleting a status update
* Streaming API

1. **Reducing memory use**

* Short structures
* Sharded structures
* Packing bits and bytes

1. **Scaling Redis**

* Scaling reads
* Scaling writes and memory capacity
* Scaling complex queries

1. **Scripting Redis with Lua**

* Adding functionality without writing C
* Rewriting locks and semaphores with Lua
* Doing away with WATCH/MULTI/EXEC
* Sharding LISTs with Lua

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with Spark 2.x

Course Snapshot

* **Course:** Machine Learning with Spark 2.x
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning with Spark skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to unlock the complexities of machine learning algorithms in Spark to generate useful data insights through this data analysis tutorial
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The purpose of machine learning is to build systems that learn from data. Being able to understand trends and patterns in complex data is critical to success; it is one of the key strategies to unlock growth in the challenging contemporary marketplace today. With the meteoric rise of machine learning, developers are now keen on finding out how can they make their Spark applications smarter.

This course gives you access to transform data into actionable knowledge. The course commences by defining machine learning primitives by the MLlib and H2O libraries. You will learn how to use Binary classification to detect the Higgs Boson particle in the huge amount of data produced by CERN particle collider and classify daily health activities using ensemble Methods for Multi-Class Classification. Next, you will solve a typical regression problem involving flight delay predictions and write sophisticated Spark pipelines. You will analyze Twitter data with help of the doc2vec algorithm and K-means clustering. Finally, you will build different pattern mining models using MLlib, perform complex manipulation of DataFrames using Spark and Spark SQL, and deploy your app in a Spark streaming environment.

Working in a hands-on learning environment, led by our Machine Learning with Spark expert instructor, students will learn about and explore:

* Process and analyze big data in a distributed and scalable way
* Write sophisticated Spark pipelines that incorporate elaborate extraction
* Build and use regression models to predict flight delays

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use Spark streams to cluster tweets online
* Run the PageRank algorithm to compute user influence
* Perform complex manipulation of DataFrames using Spark
* Define Spark pipelines to compose individual data transformations
* Utilize generated models for off-line/on-line prediction
* Transfer the learning from an ensemble to a simpler Neural Network
* Understand basic graph properties and important graph operations
* Use GraphFrames, an extension of DataFrames to graphs, to study graphs using an elegant query language
* Use K-means algorithm to cluster movie reviews dataset

Audience & Pre-Requisites

This course is designed for developers interested to unlock the complexities of machine learning algorithms in Spark to generate useful data insights through this data analysis tutorial

**Pre-Requisites:** Students should have familiar with

* Basics of Python
* Knowledge of Python is assumed.

Course Agenda / Topics

1. **Introduction to Large-Scale Machine Learning and Spark**

* Introduction to Large-Scale Machine Learning and Spark
* Data science
* The sexiest role of the 21st century – data scientist?
* Introducing H2O.ai
* What's the difference between H2O and Spark's MLlib?
* Data munging
* Data science - an iterative process

1. **Detecting Dark Matter - The Higgs-Boson Particle**

* Detecting Dark Matter - The Higgs-Boson Particle
* Type I versus type II error
* Spark start and data load

1. **Ensemble Methods for Multi-Class Classification**

* Ensemble Methods for Multi-Class Classification
* Data
* Modeling goal

1. **Predicting Movie Reviews Using NLP and Spark Streaming**

* Predicting Movie Reviews Using NLP and Spark Streaming
* NLP - a brief primer
* The dataset
* Feature extraction
* Featurization - feature hashing
* Let's do some (model) training!
* Super learner

1. **Word2vec for Prediction and Clustering**

* Word2vec for Prediction and Clustering
* Motivation of word vectors
* Word2vec explained
* Doc2vec explained
* Applying word2vec and exploring our data with vectors
* Creating document vectors
* Supervised learning task

1. **Extracting Patterns from Clickstream Data**

* Extracting Patterns from Clickstream Data
* Frequent pattern mining
* Pattern mining with Spark MLlib
* Deploying a pattern mining application

1. **Graph Analytics with GraphX**

* Graph Analytics with GraphX
* Basic graph theory
* GraphX distributed graph processing engine
* Graph algorithms and applications

1. **Lending Club Loan Prediction**

* Lending Club Loan Prediction
* Motivation
* Preparation of the environment
* Data load
* Exploration – data analysis

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Solr.

Course Snapshot

* **Course:** Solr
* **Duration:** 4 days
* **Skill-level**: Foundation-level Solr skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to implement scalable search using Apache Solr
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Whether you're handling big (or small) data, managing documents, or building a website, it is important to be able to quickly search through your content and discover meaning in it. Apache Solr is your tool: a ready-to-deploy, Lucene-based, open source, full-text search engine. Solr can scale across many servers to enable real-time queries and data analytics across billions of documents.

Solr teaches you to implement scalable search using Apache Solr. This easy-to-read guide balances conceptual discussions with practical examples to show you how to implement all of Solr's core capabilities. You'll master topics like text analysis, faceted search, hit highlighting, result grouping, query suggestions, multilingual search, advanced geospatial and data operations, and relevancy tuning.

Working in a hands-on learning environment, led by our Solr expert instructor, students will learn about and explore:

* Redis expands on the key-value pattern by accepting a wide variety of data types, including hashes, strings, lists, and other structures.
* It provides lightning-fast operations on in-memory datasets, and also makes it easy to persist to disk on the fly. Plus, it's free and open source.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* How to scale Solr for big data
* Rich real-world examples
* Solr as a NoSQL data store
* Advanced multilingual, data, and relevancy tricks
* Coverage of versions through Solr 4.7

Audience & Pre-Requisites

This course is geared for attendees who want to implement scalable search using Apache Solr

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills
* basic knowledge of Java and standard database technology.
* No prior knowledge of Solr or Lucene is required.

Course Agenda / Topics

1. **Introduction to Solrfree**

* Why do I need a search engine?
* What is Solr?
* Why Solr?
* Features overview

1. **Getting to know Solr**

* Getting started
* Searching is what it’s all about
* Tour of the Solr administration console
* Adapting the example to your needs

1. **Key Solr conceptsfree**

* Searching, matching, and finding content
* Relevancy
* Precision and Recall
* Searching at scale

1. **Configuring Solr**

* Overview of solrconfig.xml
* Query request handling
* Managing searchers
* Cache management
* Remaining configuration options

1. **Indexing**

* Example microblog search application
* Designing your schema
* Defining fields in schema.xml
* Field types for structured nontext fields
* Sending documents to Solr for indexing
* Update handler
* Index management

1. **Text analysis**

* Analyzing microblog text
* Basic text analysis
* Defining a custom field type for microblog text
* Advanced text analysis

1. **Performing queries and handling results**

* The anatomy of a Solr request
* Working with query parsers
* Queries and filters
* The default query parser (Lucene query parser)
* Handling user queries (eDisMax query parser)
* Other useful query parsers
* Returning results
* Sorting results
* Debugging query results

1. **Faceted search**

* Navigating your content at a glance
* Setting up test data
* Field faceting
* Query faceting
* Range faceting
* Filtering upon faceted values
* Multiselect faceting, keys, and tags
* Beyond the basics

1. **Hit highlighting**

* Overview of hit highlighting
* How highlighting works
* Improving performance using FastVectorHighlighter
* PostingsHighlighter

1. **Query suggestions**

* Spell-check
* Autosuggesting query terms
* Suggesting document field values
* Suggesting queries based on user activity

1. **Result grouping/field collapsing**

* Result grouping vs. field collapsing
* Skipping duplicate documents
* Returning multiple documents per group
* Grouping by functions and queries
* Paging and sorting grouped results
* Grouping gotchas
* Efficient field collapsing with the Collapsing query parser

1. **Taking Solr to production**

* Developing a Solr distribution
* Deploying Solr
* Hardware and server configuration
* Data acquisition strategies
* Sharding and replication
* Solr core management
* Managing clusters of servers
* Querying and interacting with Solr
* Monitoring Solr’s performance
* Upgrading between Solr versions

1. **SolrCloud**

* Getting started with SolrCloud
* Core concepts
* Distributed indexing
* Distributed search
* Collections API
* Basic system-administration tasks
* Advanced topics

1. **Ch 14. Multilingual search**

* Why linguistic analysis matters
* Stemming vs. lemmatization
* Stemming
* Handling edge cases
* Available language libraries in Solr
* Searching content in multiple languages
* Language identification

1. **Complex query operations**

* Function queries
* Geospatial search
* Pivot faceting
* Referencing external data
* Cross-document and cross-index joins
* Big data analytics with Solr

1. **Mastering relevancy**

* The impact of relevancy tuning
* Debugging the relevancy calculation
* Relevancy boosting
* Pluggable Similarity class implementations
* Personalized search and recommendations
* Creating a personalized search experience
* Running relevancy experiments

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Hadoop in Practice

Course Snapshot

* **Course:** Hadoop in Practice
* **Duration:** 3 days
* **Skill-level**: Foundation-level Hadoop skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to conquer big data, using Hadoop. This revised new edition covers changes and new features in the Hadoop core architecture, including MapReduce 2
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

It's always a good time to upgrade your Hadoop skills! Hadoop in Practice, This Edition provides a collection of 104 tested, instantly useful techniques for analyzing real-time streams, moving data securely, machine learning, managing large-scale clusters, and taming big data using Hadoop. This completely revised edition covers changes and new features in Hadoop core, including MapReduce 2 and YARN. You'll pick up hands-on best practices for integrating Spark, Kafka, and Impala with Hadoop, and get new and updated techniques for the latest versions of Flume, Sqoop, and Mahout. In short, this is the most practical, up-to-date coverage of Hadoop available.

Readers need to know a programming language like Java and have basic familiarity with Hadoop.

Working in a hands-on learning environment, led by our Hadoop expert instructor, students will learn about and explore:

* provides over 100 tested, instantly useful techniques that will help you conquer big data, using Hadoop.
* This revised new edition covers changes and new features in the Hadoop core architecture, including MapReduce 2.
* new lessons cover YARN and integrating Kafka, Impala, and Spark SQL with Hadoop.
* You'll also get new and updated techniques for Flume, Sqoop, and Mahout, all of which have seen major new versions recently.
* this is the most practical, up-to-date coverage of Hadoop available anywhere

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Thoroughly updated for Hadoop 2
* How to write YARN applications
* Integrate real-time technologies like Storm, Impala, and Spark
* Predictive analytics using Mahout and RR

Audience & Pre-Requisites

This course is geared for attendees who want to cover YARN and integrating Kafka, Impala, and Spark SQL with Hadoop. You'll also get new and updated techniques for Flume, Sqoop, and Mahout, all of which have seen major new versions recently.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills
* assumes you've already started exploring Hadoop and want concrete advice on how to use it in production.

Course Agenda / Topics

1. **Hadoop in a heartbeatfree**

* What is Hadoop?
* Getting your hands dirty with MapReduce

1. **Introduction to YARN**

* YARN overview
* YARN and MapReduce
* YARN applications

1. **Data serialization—working with text and beyond**

* Understanding inputs and outputs in MapReduce
* Processing common serialization formats
* Big data serialization formats
* Columnar storage
* Custom file formats

1. **Organizing and optimizing data in** **HDFS**

* Data organization
* Efficient storage with compression

1. **Moving data into and out of Hadoop**

* Key elements of data movement
* Moving data into Hadoop
* Moving data out of Hadoop

1. **Applying MapReduce patterns to big data**

* Joining
* Sorting
* Sampling

1. **Utilizing data structures and algorithms at scale**

* Modeling data and solving problems with graphs
* Bloom filters
* HyperLogLog

1. **Tuning, debugging, and testing**

* Measure, measure, measure
* Tuning MapReduce
* Debugging
* Testing MapReduce jobs

1. **SQL on Hadoop**

* Hive
* Impala
* Spark SQL

1. **Writing a YARN application**

* Fundamentals of building a YARN application
* Building a YARN application to collect cluster statistics
* Additional YARN application capabilities
* YARN programming abstraction

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Apache Hadoop 3

Course Snapshot

* **Course:** Apache Hadoop 3
* **Duration:** 2 days
* **Skill-level**: Foundation-level Apache Hadoop skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who a fast paced guide that will help you learn about Apache Hadoop 3 and its ecosystem
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Apache Hadoop is a widely used distributed data platform. It enables large datasets to be efficiently processed instead of using one large computer to store and process the data. This course will get you started with the Hadoop ecosystem, and introduce you to the main technical topics, including MapReduce, YARN, and HDFS. The course begins with an overview of big data and Apache Hadoop. Then, you will set up a pseudo Hadoop development environment and a multi-node enterprise Hadoop cluster. You will see how the parallel programming paradigm, such as MapReduce, can solve many complex data processing problems. The course also covers the important aspects of the big data software development lifecycle, including quality assurance and control, performance, administration, and monitoring. You will then learn about the Hadoop ecosystem, and tools such as Kafka, Sqoop, Flume, Pig, Hive, and HBase. Finally, you will look at advanced topics, including real time streaming using Apache Storm, and data analytics using Apache Spark. By the end of the course, you will be well versed with different configurations of the Hadoop 3 cluster.

Working in a hands-on learning environment, led by our Apache Hadoop 3 expert instructor, students will learn about and explore:

* Set up, configure and get started with Hadoop to get useful insights from large data sets
* Work with the different components of Hadoop such as MapReduce, HDFS and YARN
* Learn about the new features introduced in Hadoop 3

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Store and analyze data at scale using HDFS, MapReduce and YARN
* Install and configure Hadoop 3 in different modes
* Use Yarn effectively to run different applications on Hadoop based platform
* Understand and monitor how Hadoop cluster is managed
* Consume streaming data using Storm, and then analyze it using Spark
* Explore Apache Hadoop ecosystem components, such as Flume, Sqoop, HBase, Hive, and Kafka

Audience & Pre-Requisites

This course is geared for attendees who want a fast-paced guide that will help you learn about Apache Hadoop 3 and its ecosystem

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills

Course Agenda / Topics

1. **Hadoop 3.0 - Background and Introduction**

* Hadoop 3.0 - Background and Introduction
* How it all started
* What Hadoop is and why it is important
* How Apache Hadoop works
* Hadoop 3.0 releases and new features
* Choosing the right Hadoop distribution

1. **Planning and Setting Up Hadoop Clusters**

* Planning and Setting Up Hadoop Clusters
* Technical requirements
* Prerequisites for Hadoop setup
* Running Hadoop in standalone mode
* Setting up a pseudo Hadoop cluster
* Planning and sizing clusters
* Setting up Hadoop in cluster mode
* Diagnosing the Hadoop cluster

1. **Deep Dive into the Hadoop Distributed File System**

* Deep Dive into the Hadoop Distributed File System
* Technical requirements
* How HDFS works
* Key features of HDFS
* Data flow patterns of HDFS
* HDFS configuration files
* Hadoop filesystem CLIs
* Working with data structures in HDFS

1. **Developing MapReduce Applications**

* Developing MapReduce Applications
* Technical requirements
* How MapReduce works
* Configuring a MapReduce environment
* Understanding Hadoop APIs and packages
* Setting up a MapReduce project
* Deep diving into MapReduce APIs
* Compiling and running MapReduce jobs
* Streaming in MapReduce programming

1. **Building Rich YARN Applications**

* Building Rich YARN Applications
* Technical requirements
* Understanding YARN architecture
* Key features of YARN
* Configuring the YARN environment in a cluster
* Working with YARN distributed CLI
* Deep dive with YARN application framework
* Building and monitoring a YARN application on a cluster

1. **Monitoring and Administration of a Hadoop Cluster**

* Monitoring and Administration of a Hadoop Cluster
* Roles and responsibilities of Hadoop administrators
* Planning your distributed cluster
* Resource management in Hadoop
* High availability of Hadoop
* Securing Hadoop clusters
* Performing routine tasks

1. **Demystifying Hadoop Ecosystem Components**

* Demystifying Hadoop Ecosystem Components
* Technical requirements
* Understanding Hadoop's Ecosystem
* Working with Apache Kafka
* Writing Apache Pig scripts
* Transferring data with Sqoop
* Writing Flume jobs
* Understanding Hive
* Using HBase for NoSQL storage

1. **Advanced Topics in Apache Hadoop**

* Advanced Topics in Apache Hadoop
* Technical requirements
* Hadoop use cases in industries
* Advanced Hadoop data storage file formats
* Real-time streaming with Apache Storm
* Data analytics with Apache Spark

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Apache Hive Essentials

Course Snapshot

* **Course:** Apache Hive Essentials
* **Duration:** 3 days
* **Skill-level**: Foundation-level Apache Hive skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants a fantastic journey to discover the attributes of big data using Apache Hive
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

In this course, we prepare you for your journey into big data by frstly introducing you to backgrounds in the big data domain, alongwith the process of setting up and getting familiar with your Hive working environment. Next, the course guides you through discovering and transforming the values of big data with the help of examples. It also hones your skills in using the Hive language in an effcient manner. Toward the end, the course focuses on advanced topics, such as performance, security, and extensions in Hive, which will guide you on exciting adventures on this worthwhile big data journey. By the end of the course, you will be familiar with Hive and able to work effeciently to find solutions to big data problems

Working in a hands-on learning environment, led by our Apache Hive expert instructor, students will learn about and explore:

* Grasp the skills needed to write efficient Hive queries to analyze the Big Data
* Discover how Hive can coexist and work with other tools within the Hadoop ecosystem
* Uses practical, example-oriented scenarios to cover all the newly released features of Apache Hive 2.3.3

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Create and set up the Hive environment
* Discover how to use Hive's definition language to describe data
* Discover interesting data by joining and filtering datasets in Hive
* Transform data by using Hive sorting, ordering, and functions
* Aggregate and sample data in different ways
* Boost Hive query performance and enhance data security in Hive
* Customize Hive to your needs by using user-defined functions and integrate it
* with other tools

Audience & Pre-Requisites

This course is geared for attendees who wants a fantastic journey to discover the attributes of big data using Apache Hive

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills. Attendees without a programming background like Python may view labs as follow along exercises or team with others to complete them.
* Good foundational mathematics or logic skills

Course Agenda / Topics

1. **Overview of Big Data and Hive**

* Overview of Big Data and Hive
* A short history
* Introducing big data
* The relational and NoSQL databases versus Hadoop
* Batch, real-time, and stream processing
* Overview of the Hadoop ecosystem
* Hive overview

1. **Setting Up the Hive Environment**

* Setting Up the Hive Environment
* Installing Hive from Apache
* Installing Hive from vendors
* Using Hive in the cloud
* Using the Hive command
* Using the Hive IDE

1. **Data Definition and Description**

* Data Definition and Description
* Understanding data types
* Data type conversions
* Data Definition Language
* Database
* Tables
* Partitions
* Buckets
* Views

1. **Data Correlation and Scope**

* Data Correlation and Scope
* Project data with SELECT
* Filtering data with conditions
* Linking data with JOIN
* Combining data with UNION

1. **Data Manipulation**

* Data Manipulation
* Data exchanging with LOAD
* Data exchange with INSERT
* Data exchange with [EX|IM]PORT
* Data sorting
* Functions
* Transactions and locks

1. **Data Aggregation and Sampling**

* Data Aggregation and Sampling
* Basic aggregation
* Enhanced aggregation
* Aggregation condition
* Window functions
* Sampling

1. **Performance Considerations**

* Performance Considerations
* Performance utilities
* Design optimization
* Data optimization
* Job optimization

1. **Extensibility Considerations**

* Extensibility Considerations
* User-defined functions
* HPL/SQL
* Streaming
* SerDe

1. **Security Considerations**

* Security Considerations
* Authentication
* Authorization
* Mask and encryption

1. **Working with Other Tools**

* Working with Other Tools
* The JDBC/ODBC connector
* NoSQL
* The Hue/Ambari Hive view
* HCatalog
* Oozie
* Spark
* Hivemall

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

**Storm Applied.**

**Course Snapshot**

* **Course:** Storm Applied
* **Duration:** 2 days
* **Skill-level**: Foundation-level Storm Applied skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to use Apache Storm for the real-world tasks associated with processing and analyzing real-time data streams
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Storm Applied** is an example-driven guide to processing and analyzing real-time data streams. This immediately useful course starts by teaching you how to design Storm solutions the right way. Then, it quickly dives into real-world case studies that show you how to scale a high-throughput stream processor, ensure smooth operation within a production cluster, and more. Along the way, you'll learn to use Trident for stateful stream processing, along with other tools from the Storm ecosystem.

Working in a hands-on learning environment, led by our Apache Storm expert instructor, students will learn about and explore:

you learn how to think about designing Storm solutions the right way from day one.

But it quickly dives into real-world case studies that will bring the novice up to speed with productionizing Storm.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Mapping real problems to Storm components
* Performance tuning and scaling
* Practical troubleshooting and debugging
* Exactly-once processing with Trident

**Audience & Pre-Requisites**

This course is geared for attendees who want to build a solid foundation of Storm essentials.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* While prior experience with Storm is not assumed
* some experience with big data and real-time systems is helpful.

**Course Agenda / Topics**

1. **Introducing Stormfree**

* What is big data?
* How Storm fits into the big data picture
* Why you’d want to use Storm

1. **Core Storm conceptsfree**

* Problem definition: GitHub commit count dashboard
* Basic Storm concepts
* Implementing a GitHub commit count dashboard in Storm

1. **Topology design**

* Approaching topology design
* Problem definition: a social heat map
* Precepts for mapping the solution to Storm
* Initial implementation of the design
* Scaling the topology
* Topology design paradigms

1. **Creating robust topologies**

* Requirements for reliability
* Problem definition: a credit card authorization system
* Basic implementation of the bolts
* Guaranteed message processing
* Replay semantics

1. **Moving from local to remote topologiesfree**

* The Storm cluster
* Fail-fast philosophy for fault tolerance within a Storm cluster
* Installing a Storm cluster
* Getting your topology to run on a Storm cluster
* The Storm UI and its role in the Storm cluster

1. **Tuning in Storm**

* Problem definition: Daily Deals! reborn
* Initial implementation
* Tuning: I wanna go fast
* Latency: when external systems take their time
* Storm’s metrics-collecting API

1. **Resource contention**

* Changing the number of worker processes running on a worker node
* Changing the amount of memory allocated to worker processes (JVMs)
* Figuring out which worker nodes/processes a topology is executing on
* Contention for worker processes in a Storm cluster
* Memory contention within a worker process (JVM)
* Memory contention on a worker node
* Worker node CPU contention
* Worker node I/O contention

1. **Storm internals**

* The commit count topology revisited
* Diving into the details of an executor
* Routing and tasks
* Knowing when Storm’s internal queues overflow
* Addressing internal Storm buffers overflowing
* Tweaking buffer sizes for performance gain

1. **Trident**

* What is Trident?
* Kafka and its role with Trident
* Problem definition: Internet radio
* Implementing the internet radio design as a Trident topology
* Accessing the persisted counts through DRPC
* Mapping Trident operations to Storm primitives
* Scaling a Trident topology

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

**Big Data**

**Course Snapshot**

* **Course:** Big Data
* **Duration:** 5 days
* **Skill-level**: Foundation-level Big Data skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to use an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyze web-scale data
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Web-scale applications like social networks, real-time analytics, or e-commerce sites deal with a lot of data, whose volume and velocity exceed the limits of traditional database systems. These applications require architectures built around clusters of machines to store and process data of any size, or speed. Fortunately, scale and simplicity are not mutually exclusive.

Big Data teaches you to build big data systems using an architecture designed specifically to capture and analyze web-scale data. This course presents the Lambda Architecture, a scalable, easy-to-understand approach that can be built and run by a small team. You'll explore the theory of big data systems and how to implement them in practice. In addition to discovering a general framework for processing big data, you'll learn specific technologies like Hadoop, Storm, and NoSQL databases.

Working in a hands-on learning environment, led by our Big Data expert instructor, students will learn about and explore:

* using an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyze web-scale data.
* It describes a scalable, easy-to-understand approach to big data systems that can be built and run by a small team. this
* guides readers through the theory of big data systems, how to implement them in practice
* how to deploy and operate them once they're built.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Introduction to big data systems
* Real-time processing of web-scale data
* Tools like Hadoop, Cassandra, and Storm
* Extensions to traditional database skills

**Audience & Pre-Requisites**

This course is geared for attendees who want to use an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyze web-scale data

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* no previous exposure to large-scale data analysis or NoSQL tools.
* Familiarity with traditional databases is helpful.

**Course Agenda / Topics**

1. **A new paradigm for Big Datafree audio**
2. **Data model for Big Datafree audio**

* The properties of data
* The fact-based model for representing data
* Graph schemas
* A complete data model for SuperWebAnalytics.com

1. **Data model for Big Data: Illustration**

* Why a serialization framework?
* Apache Thrift
* Limitations of serialization frameworks

1. **Data storage on the batch layer**

* Storage requirements for the master dataset
* Choosing a storage solution for the batch layer
* How distributed filesystems work
* Storing a master dataset with a distributed filesystem
* Vertical partitioning
* Low-level nature of distributed filesystems
* Storing the SuperWebAnalytics.com master dataset on a distributed filesystem

1. **Data storage on the batch layer: Illustration**

* Using the Hadoop Distributed File System
* Data storage in the batch layer with Pail
* Storing the master dataset for SuperWebAnalytics.com

1. **Batch layer**

* Motivating examples
* Computing on the batch layer
* Recomputation algorithms vs. incremental algorithms
* Scalability in the batch layer
* MapReduce: a paradigm for Big Data computing
* Low-level nature of MapReduce
* Pipe diagrams: a higher-level way of thinking about batch computation

1. **Batch layer: Illustration**

* An illustrative example
* Common pitfalls of data-processing tools
* An introduction to JCascalog
* Composition

1. **An example batch layer: Architecture and algorithms**

* Design of the SuperWebAnalytics.com batch layer
* Workflow overview
* Ingesting new data
* URL normalization
* User-identifier normalization
* Deduplicate pageviews
* Computing batch views

1. **An example batch layer: Implementation**

* Starting point
* Preparing the workflow
* Ingesting new data
* URL normalization
* User-identifier normalization
* Deduplicate pageviews
* Computing batch views

1. **Serving layer**

* Performance metrics for the serving layer
* The serving layer solution to the normalization/denormalization problem
* Requirements for a serving layer database
* Designing a serving layer for SuperWebAnalytics.com
* Contrasting with a fully incremental solution

1. **Serving layer: Illustration**

* Basics of ElephantDB
* Building the serving layer for SuperWebAnalytics.com

1. **Realtime views**

* Computing realtime views
* Storing realtime views
* Challenges of incremental computation
* Asynchronous versus synchronous updates
* Expiring realtime views

1. **Realtime views: Illustration**

* Cassandra’s data model
* Using Cassandra

1. **Queuing and stream processing**

* Queuing
* Stream processing
* Higher-level, one-at-a-time stream processing
* SuperWebAnalytics.com speed layer

1. **Queuing and stream processing: Illustration**

* Defining topologies with Apache Storm
* Apache Storm clusters and deployment
* Guaranteeing message processing
* Implementing the SuperWebAnalytics.com uniques-over-time speed layer

1. **Micro-batch stream processing**

* Achieving exactly-once semantics
* Core concepts of micro-batch stream processing
* Extending pipe diagrams for micro-batch processing
* Finishing the speed layer for SuperWebAnalytics.com
* Another look at the bounce-rate-analysis example

1. **Micro-batch stream processing: Illustration**

* Using Trident
* Finishing the SuperWebAnalytics.com speed layer
* Fully fault-tolerant, in-memory, micro-batch processing

1. **Lambda Architecture in depth**

* Defining data systems
* Batch and serving layers
* Speed layer
* Query layer

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

**Apache Spark**

**Course Snapshot**

* **Course:** Apache Spark
* **Duration:** 2 days
* **Skill-level**: Foundation-level Apache Spark skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants a practical guide for solving complex data processing challenges by applying the best optimizations techniques in Apache Spark.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Apache Spark is a ﬂexible framework that allows processing of batch and real-time data. Its unified engine has made it quite popular for big data use cases. This course will help you to get started with Apache Spark 2.0 and write big data applications for a variety of use cases. It will also introduce you to Apache Spark – one of the most popular Big Data processing frameworks. Although this course is intended to help you get started with Apache Spark, but it also focuses on explaining the core concepts. This practical guide provides a quick start to the Spark 2.0 architecture and its components. It teaches you how to set up Spark on your local machine. As we move ahead, you will be introduced to resilient distributed datasets (RDDs) and DataFrame APIs, and their corresponding transformations and actions. Then, we move on to the life cycle of a Spark application and learn about the techniques used to debug slow-running applications. You will also go through Spark’s built-in modules for SQL, streaming, machine learning, and graph analysis. Finally, the course will lay out the best practices and optimization techniques that are key for writing efficient Spark applications. By the end of this course, you will have a sound fundamental understanding of the Apache Spark framework and you will be able to write and optimize Spark applications.

Working in a hands-on learning environment, led by our Apache Spark expert instructor, students will learn about and explore:

* Learn about the core concepts and the latest developments in Apache Spark
* Master writing efficient big data applications with Spark’s built-in modules for SQL, Streaming, Machine Learning and Graph analysis
* Get introduced to a variety of optimizations based on the actual experience

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn core concepts such as RDDs, DataFrames, transformations, and more
* Set up a Spark development environment
* Choose the right APIs for your applications
* Understand Spark’s architecture and the execution ﬂow of a Spark application
* Explore built-in modules for SQL, streaming, ML, and graph analysis
* Optimize your Spark job for better performance

**Audience & Pre-Requisites**

This course is geared for attendees who want a practical guide for solving complex data processing challenges by applying the best optimizations techniques in Apache Spark.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* no previous exposure to large-scale data analysis or NoSQL tools.
* Familiarity with traditional databases is helpful.

**Course Agenda / Topics**

1. **Introduction to Apache Spark**

* Introduction to Apache Spark
* What is Spark?
* Spark architecture overview
* Spark language APIs
* Spark components
* Making the most of Hadoop and Spark

1. **Apache Spark Installation**

* Apache Spark Installation
* AWS elastic compute cloud (EC2)
* Configuring Spark

1. **Spark RDD**

* Spark RDD
* What is an RDD?
* Programming using RDDs
* Transformations and actions
* Types of RDDs
* Caching and checkpointing
* Understanding partitions
* Drawbacks of using RDDs

1. **Spark DataFrame and Dataset**

* Spark DataFrame and Dataset
* DataFrames
* Datasets

1. **Spark Architecture and Application Execution Flow**

* Spark Architecture and Application Execution Flow
* A sample application
* Application execution modes
* Application monitoring

1. **Spark SQL**

* Spark SQL
* Spark SQL

1. **Spark Streaming, Machine Learning, and Graph Analysis**

* Spark Streaming, Machine Learning, and Graph Analysis
* Spark Streaming
* Machine learning
* Graph processing

1. **Spark Optimizations**

* Spark Optimizations
* Cluster-level optimizations
* Application optimizations

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

**Building Data Streaming Applications with Apache Kafka**

**Course Snapshot**

* **Course:** Building Data Streaming Applications with Apache Kafka
* **Duration:** 3 days
* **Skill-level**: Foundation-level Apache Spark skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Design and administer fast, reliable enterprise messaging systems with Apache Kafka
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Apache Kafka is a popular distributed streaming platform that acts as a messaging queue or an enterprise messaging system. It lets you publish and subscribe to a stream of records, and process them in a fault-tolerant way as they occur. This course is a comprehensive guide to designing and architecting enterprise-grade streaming applications using Apache Kafka and other big data tools. It includes best practices for building such applications, and tackles some common challenges such as how to use Kafka efficiently and handle high data volumes with ease. This course first takes you through understanding the type messaging system and then provides a thorough introduction to Apache Kafka and its internal details. The second part of the book takes you through designing streaming application using various frameworks and tools such as Apache Spark, Apache Storm, and more. Once you grasp the basics, we will take you through more advanced concepts in Apache Kafka such as capacity planning and security. By the end of this course, you will have all the information you need to be comfortable with using Apache Kafka, and to design efficient streaming data applications with it.

Working in a hands-on learning environment, led by our Apache Spark expert instructor, students will learn about and explore:

* Build efficient real-time streaming applications in Apache Kafka to process data streams of data
* Master the core Kafka APIs to set up Apache Kafka clusters and start writing message producers and consumers
* A comprehensive guide to help you get a solid grasp of the Apache Kafka concepts in Apache Kafka with pracitcalpractical examples

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Learn the basics of Apache Kafka from scratch
* Use the basic building blocks of a streaming application
* Design effective streaming applications with Kafka using Spark, Storm &, and Heron
* Understand the importance of a low -latency , high- throughput, and fault-tolerant messaging system
* Make effective capacity planning while deploying your Kafka Application
* Understand and implement the best security practices

**Audience & Pre-Requisites**

This course is geared for attendees who want to Design and administer fast, reliable enterprise messaging systems with Apache Kafka.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* no previous exposure to large-scale data analysis or NoSQL tools.
* Familiarity with traditional databases is helpful.

**Course Agenda / Topics**

1. **Introduction to Messaging Systems**

* Introduction to Messaging Systems
* Understanding the principles of messaging systems
* Understanding messaging systems
* Peeking into a point-to-point messaging system
* Publish-subscribe messaging system
* Advance Queuing Messaging Protocol
* Using messaging systems in big data streaming applications

1. **Introducing Kafka the Distributed Messaging Platform**

* Introducing Kafka the Distributed Messaging Platform
* Kafka origins
* Kafka's architecture
* Message topics
* Message partitions
* Replication and replicated logs
* Message producers
* Message consumers
* Role of Zookeeper

1. **Deep Dive into Kafka Producers**

* Deep Dive into Kafka Producers
* Kafka producer internals
* Kafka Producer APIs
* Java Kafka producer example
* Common messaging publishing patterns
* Best practices

1. **Deep Dive into Kafka Consumers**

* Deep Dive into Kafka Consumers
* Kafka consumer internals
* Kafka consumer APIs
* Java Kafka consumer
* Scala Kafka consumer
* Common message consuming patterns
* Best practices

1. **Building Spark Streaming Applications with Kafka**

* Building Spark Streaming Applications with Kafka
* Introduction to Spark
* Spark Streaming
* Use case log processing - fraud IP detection
* Producer

1. **Building Storm Applications with Kafka**

* Building Storm Applications with Kafka
* Introduction to Apache Storm
* Introduction to Apache Heron
* Integrating Apache Kafka with Apache Storm - Java
* Integrating Apache Kafka with Apache Storm - Scala
* Use case – log processing in Storm, Kafka, Hive

1. **Using Kafka with Confluent Platform**

* Using Kafka with Confluent Platform
* Introduction to Confluent Platform
* Deep driving into Confluent architecture
* Understanding Kafka Connect and Kafka Stream
* Playing with Avro using Schema Registry
* Moving Kafka data to HDFS

1. **Building ETL Pipelines Using Kafka**

* Building ETL Pipelines Using Kafka
* Considerations for using Kafka in ETL pipelines
* Introducing Kafka Connect
* Deep dive into Kafka Connect
* Introductory examples of using Kafka Connect
* Kafka Connect common use cases

1. **Building Streaming Applications Using Kafka Streams**

* Building Streaming Applications Using Kafka Streams
* Introduction to Kafka Streams
* Kafka Stream architecture
* Integrated framework advantages
* Understanding tables and Streams together
* Use case example of Kafka Streams

1. **Kafka Cluster Deployment**

* Kafka Cluster Deployment
* Kafka cluster internals
* Capacity planning
* Single cluster deployment
* Multicluster deployment
* Decommissioning brokers
* Data migration

1. **Using Kafka in Big Data Applications**

* Using Kafka in Big Data Applications
* Managing high volumes in Kafka
* Kafka message delivery semantics
* Big data and Kafka common usage patterns
* Kafka and data governance
* Alerting and monitoring
* Useful Kafka matrices

1. **Securing Kafka**

* Securing Kafka
* An overview of securing Kafka
* Wire encryption using SSL
* Kerberos SASL for authentication
* Understanding ACL and authorization
* Understanding Zookeeper authentication
* Apache Ranger for authorization

1. **Streaming Application Design Considerations**

* Streaming Application Design Considerations
* Latency and throughput
* Data and state persistence
* Data sources
* External data lookups
* Data formats
* Data serialization
* Level of parallelism
* Out-of-order events
* Message processing semantics

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Spark GraphX.

Course Snapshot

* **Course:** Spark GraphX
* **Duration:** 3 days
* **Skill-level**: Foundation-level Spark GraphX skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to know how to configure GraphX and how to use it interactively
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Spark GraphX** begins with the big picture of what graphs can be used for. This example-based tutorial teaches you how to use GraphX interactively. You?ll start with a crystal-clear introduction to building big data graphs from regular data, and then explore the problems and possibilities of implementing graph algorithms and architecting graph processing pipelines. Along the way, you?ll collect practical techniques for enhancing applications and applying machine learning algorithms to graph data.

Working in a hands-on learning environment, led by our GraphX expert instructor, students will learn about and explore:

* GraphX is a powerful graph processing API for the Apache Spark analytics engine that lets you draw insights from large datasets.
* GraphX gives you unprecedented speed and capacity for running massively parallel and machine learning algorithms.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understanding graph technology
* Using the GraphX API
* Developing algorithms for big graphs
* Machine learning with graphs
* Graph visualization

Audience & Pre-Requisites

This course is geared for attendees who want to collect practical techniques for enhancing applications and applying machine learning algorithms to graph data.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills
* Readers should be comfortable writing code.
* Experience with Apache Spark and Scala is not required.

Course Agenda / Topics

1. **Two important technologies: Spark and graphsfree**

* Spark: the step beyond Hadoop MapReduce
* Graphs: finding meaning from relationships
* Putting them together for lightning fast graph processing: Spark GraphX

1. **GraphX quick start**

* Getting set up and getting data
* Interactive GraphX querying using the Spark Shell
* PageRank example

1. **Some fundamentals**

* Scala, the native language of Spark
* Spark
* Graph terminology

1. **GraphX Basics**

* Vertex and edge classes
* Mapping operations
* Serialization/deserialization
* Graph generation
* Pregel API

1. **Built-in algorithms**

* Seek out authoritative nodes: PageRank
* Measuring connectedness: Triangle Count
* Find the fewest hops: ShortestPaths
* Finding isolated populations: Connected Components
* Community detection: LabelPropagation

1. **Other useful graph algorithmsfree**

* Your own GPS: Shortest Paths with Weights
* Travelling Salesman: greedy algorithm
* Route utilities: Minimum Spanning Trees

1. **Machine learning**

* Supervised, unsupervised, and semi-supervised learning
* Recommend a movie: SVDPlusPlus
* Using GraphX With MLlib
* Poor man’s training data: graph-based semi-supervised learning

1. **The missing algorithms**

* Missing basic graph operations
* Reading RDF graph files
* Poor man’s graph isomorphism: finding missing Wikipedia infobox items
* Global clustering coefficient: compare connectedness

1. **Performance and monitoring**

* Monitoring your Spark application
* Configuring Spark
* Spark performance tuning
* Graph partitioning

1. **Other languages and tools**

* Using languages other than Scala with GraphX
* Another visualization tool: Apache Zeppelin plus d3.js
* Almost a database: Spark Job Server
* Using SQL with Spark graphs with GraphFrames

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

**Streaming Data**

**Course Snapshot**

* **Course:** Streaming Data
* **Duration:** 2 days
* **Skill-level**: Foundation-level Streaming Data skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to know the concepts and requirements of streaming and real-time data systems.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Streaming Data** is an idea-rich tutorial that teaches you to think about efficiently interacting with fast-flowing data. Through relevant examples and illustrated use cases, you'll explore designs for applications that read, analyze, share, and store streaming data. Along the way, you'll discover the roles of key technologies like Spark, Storm, Kafka, Flink, RabbitMQ, and more. This course offers the perfect balance between big-picture thinking and implementation details

Working in a hands-on learning environment, led by our Streaming Data expert instructor, students will learn about and explore:

* introduces the concepts and requirements of streaming and real-time data systems.
* teaches you to think about how to efficiently interact with fast-flowing data.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* The right way to collect real-time data
* Architecting a streaming pipeline
* Analyzing the data
* Which technologies to use and when

**Audience & Pre-Requisites**

This course is geared for developers familiar with relational database concepts

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* No experience with streaming or real-time applications required.

**Course Agenda / Topics**

1. **Introducing streaming data**

* What is a real-time system?
* Differences between real-time and streaming systems
* The architectural blueprint
* Security for streaming systems
* How do we scale?

1. **Getting data from clients: data ingestion**

* Common interaction patterns
* Scaling the interaction patterns
* Fault tolerance
* A dose of reality

1. **Transporting the data from collection tier: decoupling the data pipeline**

* Why we need a message queuing tier
* Core concepts
* Security
* Fault tolerance
* Applying the core concepts to business problems

1. **Analyzing streaming data**

* Understanding in-flight data analysis
* Distributed stream-processing architecture
* Key features of stream-processing frameworks

1. **Algorithms for data analysis**

* Accepting constraints and relaxing
* Thinking about time
* Summarization techniques

1. **Storing the analyzed or collected data**

* When you need long-term storage
* Keeping it in-memory
* Use case exercises

1. **Making the data available**

* Communications patterns
* Protocols to use to send data to the client
* Filtering the stream
* Use case: building a Meetup RSVP streaming API

1. **Consumer device capabilities and limitations accessing the data**

* The core concepts
* Making it real: SuperMediaMarkets
* Introducing the web client
* The move toward a query language

1. **Analyzing Meetup RSVPs in real time**

* The collection tier
* Message queuing tier
* Analysis tier
* In-memory data store
* Data access tier

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.



CATEGORY DevOps

# Elasticsearch

Course Snapshot

* **Course:** Elasticsearch
* **Duration:** 3 days
* **Skill-level**: Foundation-level Elasticsearch skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to build scalable search applications using Elasticsearch.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Elasticsearch teaches you how to write applications that deliver professional quality search. As you read, you’ll learn to add basic search features to any application, enhance search results with predictive analysis and relevancy ranking, and use saved data from prior searches to give users a custom experience. This practical course focuses on Elasticsearch’s REST API via HTTP. Code snippets are written mostly in bash using cURL, so they’re easily translatable to other languages

Working in a hands-on learning environment, led by our Elasticsearch expert instructor, students will learn about and explore:

* teaches you how to build scalable search applications using Elasticsearch.
* You'll ramp up fast, with an informative overview and an engaging introductory example.
* you'll pick up the core concepts you need to implement basic searches and efficient indexing.
* With the fundamentals well in hand, you'll go on to gain an organized view of how to optimize your design.

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* What is a great search application?
* Building scalable search solutions
* Using Elasticsearch with any language
* Configuration and tuning

Audience & Pre-Requisites

This course is designed for developers and administrators building and managing search-oriented applications

**Pre-Requisites:** Students should have familiar with

* Basics of Programming

Course Agenda / Topics

1. [**Introducing Elasticsearch**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-1)

* [Solving search problems with Elasticsearch](https://livebook.manning.com/book/elasticsearch-in-action/chapter-1/13)
* [Exploring typical Elasticsearch use cases](https://livebook.manning.com/book/elasticsearch-in-action/chapter-1/46)

1. [**Diving into the functionality**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-2)

* [Understanding the logical layout: documents, types, and indices](https://livebook.manning.com/book/elasticsearch-in-action/chapter-2/17)
* [Understanding the physical layout: nodes and shards](https://livebook.manning.com/book/elasticsearch-in-action/chapter-2/54)
* [Indexing new data](https://livebook.manning.com/book/elasticsearch-in-action/chapter-2/98)
* [Searching for and retrieving data](https://livebook.manning.com/book/elasticsearch-in-action/chapter-2/155)
* [Configuring Elasticsearch](https://livebook.manning.com/book/elasticsearch-in-action/chapter-2/250)
* [Adding nodes to the cluster](https://livebook.manning.com/book/elasticsearch-in-action/chapter-2/295)

1. [**Indexing, updating, and deleting data**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-3)

* [Using mappings to define kinds of documents](https://livebook.manning.com/book/elasticsearch-in-action/chapter-3/15)
* [Core types for defining your own fields in documents](https://livebook.manning.com/book/elasticsearch-in-action/chapter-3/46)
* [Arrays and multi-fields](https://livebook.manning.com/book/elasticsearch-in-action/chapter-3/95)
* [Using predefined fields](https://livebook.manning.com/book/elasticsearch-in-action/chapter-3/112)
* [Updating existing documents](https://livebook.manning.com/book/elasticsearch-in-action/chapter-3/178)
* [Deleting data](https://livebook.manning.com/book/elasticsearch-in-action/chapter-3/239)

1. [**Searching your data**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-4)

* [Structure of a search request](https://livebook.manning.com/book/elasticsearch-in-action/chapter-4/13)
* [Introducing the query and filter DSL](https://livebook.manning.com/book/elasticsearch-in-action/chapter-4/71)
* [Combining queries or compound queries](https://livebook.manning.com/book/elasticsearch-in-action/chapter-4/151)
* [Beyond match and filter queries](https://livebook.manning.com/book/elasticsearch-in-action/chapter-4/175)
* [Querying for field existence with filters](https://livebook.manning.com/book/elasticsearch-in-action/chapter-4/205)
* [Choosing the best query for the job](https://livebook.manning.com/book/elasticsearch-in-action/chapter-4/224)

1. [**Analyzing your data**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-5)

* [What is analysis?](https://livebook.manning.com/book/elasticsearch-in-action/chapter-5/10)
* [Using analyzers for your documents](https://livebook.manning.com/book/elasticsearch-in-action/chapter-5/31)
* [Analyzing text with the analyze API](https://livebook.manning.com/book/elasticsearch-in-action/chapter-5/60)
* [Analyzers, tokenizers, and token filters, oh my!](https://livebook.manning.com/book/elasticsearch-in-action/chapter-5/91)
* [Ngrams, edge ngrams, and shingles](https://livebook.manning.com/book/elasticsearch-in-action/chapter-5/202)
* [Stemming](https://livebook.manning.com/book/elasticsearch-in-action/chapter-5/235)

1. [**Searching with relevancy**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6)

* [How scoring works in Elasticsearch](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/13)
* [Other scoring methods](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/38)
* [Boosting](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/65)
* [Understanding how a document was scored with explain](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/91)
* [Reducing scoring impact with query rescoring](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/109)
* [Custom scoring with function\_score](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/118)
* [Tying it back together](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/181)
* [Sorting with scripts](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/195)
* [Field data detour](https://livebook.manning.com/book/elasticsearch-in-action/chapter-6/203)

1. [**Exploring your data with aggregations**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-7)

* [Understanding the anatomy of an aggregation](https://livebook.manning.com/book/elasticsearch-in-action/chapter-7/21)
* [Metrics aggregations](https://livebook.manning.com/book/elasticsearch-in-action/chapter-7/57)
* [Multi-bucket aggregations](https://livebook.manning.com/book/elasticsearch-in-action/chapter-7/107)
* [Nesting aggregations](https://livebook.manning.com/book/elasticsearch-in-action/chapter-7/192)

1. [**Relations among documents**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-8)

* [Overview of options for defining relationships among documents](https://livebook.manning.com/book/elasticsearch-in-action/chapter-8/9)
* [Having objects as field values](https://livebook.manning.com/book/elasticsearch-in-action/chapter-8/41)
* [Nested type: connecting nested documents](https://livebook.manning.com/book/elasticsearch-in-action/chapter-8/76)
* [Parent-child relationships: connecting separate documents](https://livebook.manning.com/book/elasticsearch-in-action/chapter-8/157)
* [Denormalizing: using redundant data connections](https://livebook.manning.com/book/elasticsearch-in-action/chapter-8/247)
* [Application-side joins](https://livebook.manning.com/book/elasticsearch-in-action/chapter-8/311)

1. [**Scaling out**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9)

* [Adding nodes to your Elasticsearch cluster](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9/10)
* [Discovering other Elasticsearch nodes](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9/25)
* [Removing nodes from a cluster](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9/48)
* [Upgrading Elasticsearch nodes](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9/71)
* [Using the \_cat API](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9/100)
* [Scaling strategies](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9/110)
* [Aliases](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9/130)
* [Routing](https://livebook.manning.com/book/elasticsearch-in-action/chapter-9/158)

1. [**Improving performance**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-10)

* [Grouping requests](https://livebook.manning.com/book/elasticsearch-in-action/chapter-10/15)
* [Optimizing the handling of Lucene segments](https://livebook.manning.com/book/elasticsearch-in-action/chapter-10/70)
* [Making the best use of caches](https://livebook.manning.com/book/elasticsearch-in-action/chapter-10/166)
* [Other performance tradeoffs](https://livebook.manning.com/book/elasticsearch-in-action/chapter-10/273)

1. [**Administering your cluster**](https://livebook.manning.com/book/elasticsearch-in-action/chapter-11)

* [Improving defaults](https://livebook.manning.com/book/elasticsearch-in-action/chapter-11/10)
* [Allocation awareness](https://livebook.manning.com/book/elasticsearch-in-action/chapter-11/78)
* [Monitoring for bottlenecks](https://livebook.manning.com/book/elasticsearch-in-action/chapter-11/104)
* [Backing up your data](https://livebook.manning.com/book/elasticsearch-in-action/chapter-11/234)

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# AI as a Service

Course Snapshot

* **Course:** AI as a Service
* **Duration:** 2 days
* **Skill-level**: Foundation-level AI as a Service for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for software developers who wants to build and implement serverless AI applications, without bogging you down with a lot of theory. Instead, you’ll find easy-to-digest instruction and two complete hands-on serverless AI builds in this must-have guide!
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**AI as a Service** teaches you how to quickly harness the power of serverless computing and cloud-based AI services. After an introduction to the basics of this dynamic technology duo, you’ll dive right into your first hands-on serverless AI project: a system that can recognize images from arbitrary web pages. In it you’ll explore tools like Amazon Recognition for image analysis and techniques like deployment of cloud infrastructure, a crawler service, and a simple API. When you’ve mastered the concepts and skills in that fun and interesting project, you’ll move on to building a serverless to-do application that employs cloud-based AI tools like AWS Transcribe and Polly for speech-to-text and text-to-speech functionality and Lex for creating interactive chatbots. When you’re finished with this essential course, you’ll have the skills to quickly build end-to-end serverless AI systems, making you indispensable as this rapidly emerging paradigm becomes the business standard!

Working in a hands-on learning environment, led by our AI expert instructor, students will learn about and explore:

* you’ll explore tools like Amazon Recognition for image analysis and techniques like deployment of cloud infrastructure, a crawler service, and a simple API.
* you’ll move on to building a serverless to-do application that employs cloud-based AI tools like AWS Transcribe and Polly for speech-to-text and text-to-speech functionality and Lex for creating interactive chatbots

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Cloud AI from development to production
* Applying cloud AI services to your existing platform
* Understanding orchestration patterns for cloud AI systems
* How to architect and build scalable, resilient data pipelines
* Debugging and troubleshooting cloud AI services
* Getting started immediately with serverless templates

Audience & Pre-Requisites

This course is geared for those who wants to build and implement serverless AI applications, without bogging you down with a lot of theory.

**Pre-Requisites:** Students should have

* For software developers with intermediate skills in at least one programming language and a basic understanding of IP networking and HTTP protocol.
* Familiarity with cloud-based version control systems such as GitHub would be helpful.
* No prior knowledge of AI is necessary.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **A Tale of Two Technologies**

* Cloud Landscape
* What is Serverless?
* The Need for Speed
* What is AI?
* The Democratization of Compute Power and Artificial Intelligence
* Canonical Serverless AI architecture
* Realization on Amazon Web Services

1. **Building a Serverless Image Recognition System**

* Our First System
* Architecture
* Getting ready
* Implementing the Asynchronous services
* Implementing the Synchronous Services
* Running the System
* Removing the System

1. **Build and Secure a Web Application the Serverless Way**

* The To Do List
* Architecture
* Getting ready
* Step 1 The Basic Application
* Step 2 Securing with Cognito

1. **Adding AI Interfaces to a Web Application**

* Step 3 Adding a Speech to Text Interface
* Step 4 Adding Text to Speech
* Step 5 Adding a Conversational Chat Bot Interface
* Removing the System

1. **Applying AI to Existing Platforms**

* Integration Patterns for Serverless AI
* Improving identity verification with Textract
* An AI Enabled Data Processing Pipeline with Kinesis
* Deploying the API
* On the fly translation with Translate
* Testing the Pipeline
* Sentiment Analysis with Comprehend
* Training a Custom Document Classifier
* Using the Custom Classifier
* Testing the Pipeline End to End
* Removing the Pipeline
* Benefits of Automation

1. **Gathering Data at Scale for Real-World AI**

* Scenario: Finding Events and Speakers
* Gathering Data from the Web
* Introduction to Web Crawling
* Implementing an Item Store
* Creating a Frontier to Store and Manage URLs
* Building the Fetcher to Retrieve and Parse Web Pages
* Determining the Crawl Space in a Strategy Service
* Orchestrating the Crawler with a Scheduler

1. **Extracting Value from Large Data Sets with AI**

* Using AI to Extract Significant Information from Web Pages
* Understanding Comprehend Entity Recognition APIs
* Preparing Data for Information Extraction
* Managing Throughput with Text Batches
* Asynchronous Named Entity Abstraction
* Checking Entity Recognition Progress
* Deploying and Testing Batch Entity Recognition
* Persisting Recognition Results
* Tieing it all Together
* Wrapping Up

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Introduction to DevOps with Kubernetes

Course Snapshot

* **Course:** Introduction to DevOps with Kubernetes
* **Duration:** 2 days
* **Skill-level**: Foundation-level DevOps with Kubernetes for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for software developers who wants to Become familiar with Kubernetes and explore techniques to manage your containerized workloads and services
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Kubernetes and DevOps are the two pillars that can keep your business at the top by ensuring high performance of your IT infrastructure. Introduction to DevOps with Kubernetes will help you develop the skills you need to improve your DevOps with the power of Kubernetes. The course begins with an overview of Kubernetes primitives and DevOps concepts. You'll understand how Kubernetes can assist you with overcoming a wide range of real-world operation challenges. You will get to grips with creating and upgrading a cluster, and then learn how to deploy, update, and scale an application on Kubernetes. As you advance through the lessons, you’ll be able to monitor an application by setting up a pod failure alert on Prometheus. The course will also guide you in configuring Alert manager to send alerts to the Slack channel and trace down a problem on the application using kubectl commands. By the end of this course, you’ll be able to manage the lifecycle of simple to complex applications on Kubernetes with confidence.

Working in a hands-on learning environment, led by our DevOps with Kubernetes expert instructor, students will learn about and explore:

* Learn everything from creating a cluster to monitoring applications in Kubernetes
* Understand and develop DevOps primitives using Kubernetes
* Use Kubernetes to solve challenging real-life DevOps problems

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Create and manage Kubernetes clusters in on-premise systems and cloud
* Exercise various DevOps practices using Kubernetes
* Explore configuration, secret, and storage management, and exercise these on Kubernetes
* Perform different update techniques and apply them on Kubernetes
* Use the built-in scaling feature in Kubernetes to scale your applications up and down
* Use various troubleshooting techniques and have a monitoring system installed on Kubernetes

Audience & Pre-Requisites

This course is geared for those who wants to Become familiar with Kubernetes and explore techniques to manage your containerized workloads and services

**Pre-Requisites:** Students should have

* Familiarity with cloud-based version control systems such as GitHub would be helpful.
* No prior knowledge of AI is necessary.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Introduction to DevOps**

* Introduction to DevOps
* Introduction
* The DevOps Toolchain

1. **Introduction to** **Microservices and Containers**

* Introduction to Microservices and Containers
* Introduction
* Introduction to Docker
* Building Docker Images
* Running Docker Containers

1. **Introduction to Kubernetes**

* Introduction to Kubernetes
* Introduction
* Kubernetes Architecture
* Accessing Kubernetes Clusters
* Fundamental Kubernetes Resources

1. **Creating a Kubernetes Cluster**

* Creating a Kubernetes Cluster
* Introduction
* Manual Kubernetes Cluster Setup
* Kubernetes Cluster Considerations
* Kubernetes Platform Options

1. **Deploy an Application to Kubernetes**

* Deploy an Application to Kubernetes
* Introduction
* Object Management in Kubernetes
* Service Discovery in Kubernetes
* Kubernetes Package Manager: Helm

1. **Configuration and Storage** **Management in Kubernetes**

* Configuration and Storage Management in Kubernetes
* Configuration Management
* Configuration Management in Kubernetes
* Secret Management
* Secret Management Best Practices
* Storage Management
* Storage Management in Kubernetes
* Activity 7: Running a Persistent Database on Kubernetes

1. **Updating and Scaling an Application in Kubernetes**

* Updating and Scaling an Application in Kubernetes
* Updating an Application
* Updating an Application in Kubernetes
* Scaling an Application in Kubernetes

1. **Troubleshooting Applications in Kubernetes**

* Troubleshooting Applications in Kubernetes
* Introduction
* Troubleshooting
* Troubleshooting Applications in Kubernetes

1. **Monitoring Applications in Kubernetes**

* Monitoring Applications in Kubernetes
* Monitoring
* Monitoring Applications in Kubernetes
* Custom Metrics
* Exposed Metrics in Kubernetes

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Automated Machine Learning

Course Snapshot

* **Course:** Automated Machine Learning
* **Duration:** 2 days
* **Skill-level**: Foundation-level Automated Machine Learning for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for software developers who wants to Automate data and model pipelines for faster machine learning applications
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation, or remote instructor led delivery, or CBT/WBT (by request).
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

AutoML is designed to automate parts of Machine Learning. Readily available AutoML tools are making data science practitioners’ work easy and are received well in the advanced analytics community. Automated Machine Learning covers the necessary foundation needed to create automated machine learning modules and helps you get up to speed with them in the most practical way possible. In this course, you’ll learn how to automate different tasks in the machine learning pipeline such as data preprocessing, feature selection, model training, model optimization, and much more. In addition to this, it demonstrates how you can use the available automation libraries, such as auto-sklearn and MLBox, and create and extend your own custom AutoML components for Machine Learning. By the end of this course, you will have a clearer understanding of the different aspects of automated Machine Learning, and you’ll be able to incorporate automation tasks using practical datasets. You can leverage your learning from this course to implement Machine Learning in your projects and get a step closer to winning various machine learning competitions.

Working in a hands-on learning environment, led by our Automated Machine Learning expert instructor, students will learn about and explore:

* Build automated modules for different machine learning components
* Understand each component of a machine learning pipeline in depth
* Learn to use different open source AutoML and feature engineering platforms

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Understand the fundamentals of Automated Machine Learning systems
* Explore auto-sklearn and MLBox for AutoML tasks
* Automate your preprocessing methods along with feature transformation
* Enhance feature selection and generation using the Python stack
* Assemble individual components of ML into a complete AutoML framework
* Demystify hyperparameter tuning to optimize your ML models
* Dive into Machine Learning concepts such as neural networks and autoencoders
* Understand the information costs and trade-offs associated with AutoML

Audience & Pre-Requisites

This course is geared for those who wants to Automate data and model pipelines for faster machine learning applications

**Pre-Requisites:** Students should have

* For software developers with intermediate skills in at least one programming language and a basic understanding of IP networking and HTTP protocol.
* No prior knowledge of AI is necessary.
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Introduction to AutoML**

* Introduction to AutoML
* Scope of machine learning
* What is AutoML?
* Why use AutoML and how does it help?
* When do you automate ML?
* What will you learn?
* Overview of AutoML libraries

1. **Introduction to Machine Learning Using Python**

* Introduction to Machine Learning Using Python
* Technical requirements
* Machine learning
* Linear regression
* Important evaluation metrics – regression algorithms
* Logistic regression
* Important evaluation metrics – classification algorithms
* Decision trees
* Support Vector Machines
* k-Nearest Neighbors
* Ensemble methods
* Comparing the results of classifiers
* Cross-validation
* Clustering

1. **Data Preprocessing**

* Data Preprocessing
* Technical requirements
* Data transformation
* Feature selection
* Feature generation

1. **Automated Algorithm Selection**

* Automated Algorithm Selection
* Technical requirements
* Computational complexity
* Differences in training and scoring time
* Linearity versus non-linearity
* Necessary feature transformations
* Supervised ML
* Unsupervised AutoML

1. **Hyperparameter Optimization**

* Hyperparameter Optimization
* Technical requirements
* Hyperparameters
* Warm start
* Bayesian-based hyperparameter tuning
* An example system

1. **Creating AutoML Pipelines**

* Creating AutoML Pipelines
* Technical requirements
* An introduction to machine learning pipelines
* A simple pipeline
* FunctionTransformer
* A complex pipeline

1. **Dive into Deep Learning**

* Dive into Deep Learning
* Technical requirements
* Overview of neural networks
* A feed-forward neural network using Keras
* Autoencoders
* Convolutional Neural Networks

1. **Critical Aspects of ML and Data Science Projects**

* Critical Aspects of ML and Data Science Projects
* Machine learning as a search
* Trade-offs in machine learning
* Engagement model for a typical data science project
* The phases of an engagement model

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Real-World Cryptography

Course Snapshot

* **Course:** Real-World Cryptography
* **Duration:** 2 days
* **Skill-level**: Foundation-level Real-World Cryptography skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants applied cryptographic techniques to understand and apply security at every level of your systems and applications.
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Real-World Cryptography** helps you understand the cryptographic techniques at work in common tools, frameworks, and protocols so you can make excellent security choices for your systems and applications. There’s no unnecessary theory or jargon—just the most up-to-date techniques you’ll need in your day-to-day work as a developer or systems administrator. Cryptography expert David Wong takes you hands-on with cryptography building blocks such as hash functions and key exchanges, then shows you how to use them as part of your security protocols and applications. Alongside modern methods, the course also anticipates the future of cryptography, diving into emerging and cutting-edge advances such as cryptocurrencies, password-authenticated key exchange, and post-quantum cryptography. Throughout, all techniques are fully illustrated with diagrams and real-world use cases so you can easily see how to put them into practice.

Working in a hands-on learning environment, led by our Real-World Cryptography expert instructor, students will learn about and explore:

* Without cryptographic techniques allowing for easy encrypting and decrypting of data, almost all IT infrastructure would be vulnerable.
* Explore how cryptography protects privacy, secures online activity, and defends confidential information, such as credit cards, from attackers and thieves

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Best practices for using cryptography
* Diagrams and explanations of cryptographic algorithms
* Identifying and fixing cryptography bad practices in applications
* Picking the right cryptographic tool to solve problems

Audience & Pre-Requisites

This course is geared for cryptography beginners with no previous experience in the field.

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills.
* Good foundational mathematics or logic skills
* no previous experience is required in this field.
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

Course Agenda / Topics

1. **Real-World Cryptography**

* A Peek into the World of Cryptography
* Real World Cryptography
* A Word of Warning

1. **Hash Functions**

* What Is a Hash Function?
* Security Properties of a Hash Function
* Security Considerations for Hash Functions
* Hash Functions in practice
* Standardized Hash Functions
* Hashing Passwords

1. **Message Authentication Codes**

* What Is a Message Authentication Code?
* Security Properties of a Message Authentication Code
* MAC in the real world
* Message Authentication Codes in Practice
* Attack of the Ch: SHA-2 and Length-Extension Attacks

1. **Authenticated Encryption**

* What Is a Cipher?
* Symmetric Encryption in the Real World
* The AES-CBC-HMAC Encryption Algorithm
* Authenticated Encryption with Associated Data (AEAD)
* Key Wrapping and Nonce-Misuse Resistance
* A Map of Authenticated Encryption
* Other Kinds of Symmetric Encryption

1. **Key Exchanges**

* What is a Key Exchange?
* Key Exchange Standards

1. **Asymmetric Encryption and Hybrid Encryption**

* What is Asymmetric Encryption?
* Asymmetric Encryption in Practice and Hybrid Encryption
* Standards for Asymmetric Encryption and Hybrid Encryption
* Attack of the Ch: RSA PKCS#1 v1.5
* Asymmetric Encryption with RSA-OAEP
* Hybrid Encryption with ECIES

1. **Digital Signatures**

* What is a Digital Signature?
* Security Properties and Considerations
* Digital Signature Standards

1. **Randomness and Secrets**

* What is Randomness?
* What is a Pseudo-Random Number Generator (PRNG)?
* Obtaining Randomness in Practice
* Randomness Generation and Security Considerations
* Public Randomness
* Key Derivation With HKDF
* Managing Keys and Secrets
* Avoiding Key Management, Or How to Split Trust

1. **Secure Transport (Session Encryption)**

* What is SSL/TLS?
* How Does TLS Work?
* The State of the Encrypted Web Today
* Other Secure Transport Protocols

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.



CATEGORY Cloud

# Machine Learning with Azure

**Course Snapshot**

* **Course:** Machine Learning with Azure
* **Duration:** 3 days
* **Skill-level**: Foundation-level Linux and azure skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Implement machine learning, cognitive services, and artificial intelligence solutions by leveraging Azure cloud technologies
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Implementing Machine learning (ML) and Artificial Intelligence (AI)** in the cloud had not been possible earlier due to the lack of processing power and storage. However, Azure has created ML and AI services that are easy to implement in the cloud. Hands-On Machine Learning with Azure teaches you how to perform advanced ML projects in the cloud in a cost-effective way. The course begins by covering the benefits of ML and AI in the cloud. You will then explore Microsoft’s Team Data Science Process to establish a repeatable process for successful AI development and implementation. You will also gain an understanding of AI technologies available in Azure and the Cognitive Services APIs to integrate them into bot applications. This course lets you explore prebuilt templates with Azure Machine Learning Studio and build a model using canned algorithms that can be deployed as web services. The course then takes you through a preconfigured series of virtual machines in Azure targeted at AI development scenarios. You will get to grips with the ML Server and its capabilities in SQL and HDInsight. In the concluding lessons, you’ll integrate patterns with other non-AI services in Azure.

Working in a hands-on learning environment, led by our Machine Learning with Azure expert instructor, students will learn about and explore:

* Learn advanced concepts in Azure ML and the Cortana Intelligence Suite architecture
* Explore ML Server using SQL Server and HDInsight capabilities
* Implement various tools in Azure to build and deploy machine learning models

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Discover the benefits of leveraging the cloud for ML and AI
* Use Cognitive Services APIs to build intelligent bots
* Build a model using canned algorithms from Microsoft and deploy it as a web service
* Deploy virtual machines in AI development scenarios
* Apply R, Python, SQL Server, and Spark in Azure
* Build and deploy deep learning solutions with CNTK, MMLSpark, and TensorFlow
* Implement model retraining in IoT, Streaming, and Blockchain solutions
* Explore best practices for integrating ML and AI functions with ADLA and logic apps

**Audience & Pre-Requisites**

This course is geared for attendees with basic Linux and computing skills who wish to intending to start practical Machine Learning with azure

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, Microsoft azure and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **AI Cloud Foundations**

* The importance of artificial intelligence
* The emergence of the cloud
* The Microsoft cloud – Azure

1. **Data Science Process**

* TDSP stages
* Tools for TDSP

1. **Cognitive Services**

* Cognitive Services for Vision APIs
* The Computer Vision API
* Face API
* Cognitive Services for Language APIs
* Cognitive Services for Speech APIs
* Cognitive Services for Knowledge APIs
* Cognitive Services for Search APIs

1. **Bot Framework**

* What is a bot?
* Creating a bot with Bot Service

1. **Azure Machine Learning Studio**

* Deploying an Azure AI Gallery template
* Building an experiment
* Deploying a model as a web service

1. **Scalable Computing for Data Science**

* Different scalable compute options in Azure
* Introduction to DSVMs
* DLVM
* Batch AI service
* ACI
* AKS

1. **Machine Learning Server**

* What is Microsoft ML Server?
* Machine learning with Python

1. **HDInsight**

* R with HDInsight
* Getting started with Azure HDInsight and ML services
* HDInsight and data analytics with R
* Enriching data for analysis

1. **Machine Learning with Spark**

* Machine learning with Azure Databricks
* Getting started with Apache Spark and Azure Databricks
* Using SQL in Azure Databricks
* Machine Learning with HDInsight
* HDInsight and Spark
* Working with data in a Spark environment
* Configuring the data science virtual machine
* Setting up an HDInsight cluster with Spark

1. **Building Deep Learning Solutions**

* What is deep learning?
* Overview of the Azure Notebook service
* Overview of Azure Deep Learning Virtual Machine toolkits
* An overview of the Microsoft Machine Learning Library for Apache Spark (MMLSpark)
* Overview of TensorFlow on Azure

1. **Integration with Other Azure Services**

* Logic Apps
* Azure Functions
* Azure Data Lake Analytics
* Azure Data Factory

1. **End-to-End Machine Learning**

* Using the Azure Machine Learning SDK for E2E machine learning

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Cloud Analytics with Microsoft Azure

**Course Snapshot**

* **Course:** Cloud Analytics with Microsoft Azure
* **Duration:** 2 days
* **Skill-level**: Foundation-level Cloud Analytics with Microsoft Azure skills for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to get Leverage the power of Azure to get efficient data insights from your big data in real time
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

With data being generated at an exponential speed, organizations all over the world are migrating their infrastructure to the cloud. Application management becomes much easier when you use a cloud platform to build, manage, and deploy your services and applications. Cloud Analytics with Microsoft Azure covers all that you need to extract useful insights from your data. You'll explore the power of data with big data analytics, the Internet of Things (IoT), machine learning, artificial intelligence, and DataOps. You’ll also delve into data analytics by studying use cases that focus on creating actionable insights from near-real-time data. As you advance, you’ll learn to build an end-to-end analytics pipeline on the cloud with machine learning and deep learning concepts. By the end of this course, you'll have developed a solid understanding of data analytics with Azure and its practical implementation.

Working in a hands-on learning environment, led by our Cloud Analytics with Microsoft Azure instructor, students will learn about and explore:

* Explore the basics of cloud analytics using Azure
* Discover different ways to process and visualize your data easily
* Learn to use Azure Synapse Analytics (formerly known as Azure SQL Data Warehouse) to derive real-time customer insights

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

Explore the concepts of modern data warehouses and data pipelines

Discover different design considerations while applying a cloud analytics solution

Design an end-to-end analytics pipeline on the cloud

Differentiate between structured, semi-structured, and unstructured data

Choose a cloud-based service for your data analytics solutions

Use Azure services to ingest, store and analyze data of any scale

**Audience & Pre-Requisites**

This course is geared for attendees with basic Linux and computing skills who wish to Leverage the power of Azure to get efficient data insights from your big data in real time

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, Microsoft azure and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introducing Analytics on Azure**

* The Power of Data
* Big Data Analytics
* Internet of Things (IoT)
* Machine Learning and Artificial Intelligence
* DataOps
* Why Microsoft Azure?
* Top Business Drivers for Adopting Data Analytics on the Cloud
* Why Do You Need a Modern Data Warehouse?
* Creating a Data Pipeline
* Smarter Applications

1. **Building Your Modern Data Warehouse**

* What is a Modern Data Warehouse?
* Azure Synapse Analytics
* Azure Data Factory
* Azure Data Lake Storage Gen2
* Azure Databricks
* Quick Start Guide

1. **Processing and Visualizing Data**

* Azure Analysis Services
* Power BI
* Quick Start Guide (Data Modeling and Visualization)
* Machine Learning on Azure
* Azure Machine Learning Services Features and Benefits
* Quick Start Guide (Machine Learning)

1. **Introducing Azure Synapse Analytics**

* What is Azure Synapse Analytics?
* Why do we need Azure Synapse Analytics?
* The Modern Data Warehouse Pattern
* Deep Dive into Azure Synapse Analytics
* New Preview Features
* Upcoming Changes

1. **Business Use Cases**

* Use Case 1: Real-Time Customer Insights with Azure Synapse Analytics
* The Problem
* Design Brainstorming
* The Solution
* Azure Services
* Insights and Actions
* Use Case 2: Using Advanced Analytics on Azure to Create a Smart Airport
* The Problem
* Design Brainstorming
* The Solution
* Azure Services
* Insights and Actions
* Conclusion

1. **Azure Modern Data Warehouse Life Cycle**

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning on Google Cloud Platform

**Course Snapshot**

* **Course:** Machine Learning on Google Cloud Platform
* **Duration:** 4 days
* **Skill-level**: Foundation-level Machine Learning on Google Cloud Platform for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Unleash Google's Cloud Platform to build, train and optimize machine learning models
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

**Google Cloud Machine Learning Engine** combines the services of Google Cloud Platform with the power and flexibility of TensorFlow. With this course, you will not only learn to build and train different complexities of machine learning models at scale but also host them in the cloud to make predictions. This course is focused on making the most of the Google Machine Learning Platform for large datasets and complex problems. You will learn from scratch how to create powerful machine learning based applications for a wide variety of problems by leveraging different data services from the Google Cloud Platform. Applications include NLP, Speech to text, Reinforcement learning, Time series, recommender systems, image classification, video content inference and many other. We will implement a wide variety of deep learning use cases and also make extensive use of data related services comprising the Google Cloud Platform ecosystem such as Firebase, Storage APIs, Datalab and so forth. This will enable you to integrate Machine Learning and data processing features into your web and mobile applications. By the end of this course, you will know the main difficulties that you may encounter and get appropriate strategies to overcome these difficulties and build efficient systems.

Working in a hands-on learning environment, led by our Machine Learning on Google Cloud Platform expert instructor, students will learn about and explore:

* Get well versed in GCP pre-existing services to build your own smart models
* A comprehensive guide covering aspects from data processing, analyzing to building and training ML models
* A practical approach to produce your trained ML models and port them to your mobile for easy access

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use Google Cloud Platform to build data-based applications for dashboards, web, and mobile
* Create, train and optimize deep learning models for various data science problems on big data
* Learn how to leverage BigQuery to explore big datasets
* Use Google’s pre-trained TensorFlow models for NLP, image, video and much more
* Create models and architectures for Time series, Reinforcement Learning, and generative models
* Create, evaluate, and optimize TensorFlow and Keras models for a wide range of applications

**Audience & Pre-Requisites**

This course is geared for attendees who wish to Unleash Google's Cloud Platform to build, train and optimize machine learning models

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introducing the Google Cloud Platform**

* Introducing the Google Cloud Platform
* ML and the cloud
* Introducing the GCP
* Getting started with GCP

1. **Google Compute Engine**

* Google Compute Engine
* Google Compute Engine
* Setting up a data science stack on the VM
* BOX the ipython console

1. **Google Cloud Storage**

* Google Cloud Storage
* Google Cloud Storage
* Accessing control lists
* Creating a bucket in Google Cloud Storage
* Life cycle management
* Google Cloud SQL

1. **Querying Your Data with BigQuery**

* Querying Your Data with BigQuery
* Approaching big data
* Data structuring
* Querying the database
* Google BigQuery
* Visualizing data with Google Data Studio

1. **Transforming Your Data**

* Transforming Your Data
* How to clean and prepare the data
* Finding outliers in the data
* Run Job
* Scale of features
* Google Cloud Dataflow

1. **Essential Machine Learning**

* Essential Machine Learning
* Applications of machine learning
* Supervised and unsupervised machine learning
* Overview of machine learning techniques

1. **Google Machine Learning APIs**

* Google Machine Learning APIs
* Vision API
* Cloud Translation API
* Natural Language API
* Speech-to-text API
* Video Intelligence API

1. **Creating ML Applications with Firebase**

* Creating ML Applications with Firebase
* Features of Firebase

1. **Neural Networks with TensorFlow and Keras**

* Neural Networks with TensorFlow and Keras
* Overview of a neural network

1. **Evaluating Results with TensorBoard**

* Evaluating Results with TensorBoard
* Setting up TensorBoard
* Overview of summary operations

1. **Optimizing the Model through Hyperparameter Tuning**

* Optimizing the Model through Hyperparameter Tuning
* The intuition of hyperparameter tuning

1. **Preventing Overfitting with Regularization**

* Preventing Overfitting with Regularization
* Intuition of over/under fitting

1. **Beyond Feedforward Networks – CNN and RNN**

* Beyond Feedforward Networks – CNN and RNN
* Convolutional neural networks
* Handwriting Recognition using CNN and TensorFlow
* Recurrent neural network
* Long short-term memory networks
* Handwriting Recognition using RNN and TensorFlow

1. **Time Series with LSTMs**

* Time Series with LSTMs
* Introducing time series
* Classical approach to time series
* Time series models
* Removing seasonality from a time series
* LSTM for time series analysis

1. **Reinforcement Learning**

* Reinforcement Learning
* Reinforcement learning introduction
* Reinforcement learning techniques
* OpenAI Gym
* Cart-Pole system

1. **Generative Neural Networks**

* Generative Neural Networks
* Unsupervised learning
* Generative models
* Feature extraction using RBM
* Autoencoder with Keras
* Magenta

1. **Chatbots**

* Chatbots
* Chatbots fundamentals
* Google Cloud Dialogflow

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Machine Learning with AWS

**Course Snapshot**

* **Course:** Machine Learning with AWS
* **Duration:** 2 days
* **Skill-level**: Foundation-level Machine Learning for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to use artificial intelligence and machine learning on AWS to create engaging applications
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Machine Learning with AWS is the right place to start if you are a beginner interested in learning useful artificial intelligence (AI) and machine learning skills using Amazon Web Services (AWS), the most popular and powerful cloud platform. You will learn how to use AWS to transform your projects into apps that work at high speed and are highly scalable. From natural language processing (NLP) applications, such as language translation and understanding news articles and other text sources, to creating chatbots with both voice and text interfaces, you will learn all that there is to know about using AWS to your advantage. You will also understand how to process huge numbers of images fast and create machine learning models. By the end of this course, you will have developed the skills you need to efficiently use AWS in your machine learning and artificial intelligence projects.

Working in a hands-on learning environment, led by our Machine Learning expert instructor, students will learn about and explore:

* Explore popular AI and ML services with their underlying algorithms
* Use the AWS environment to manage your AI workflow
* Reinforce key concepts with hands-on exercises using real-world datasets

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Get up and running with machine learning on the AWS platform
* Analyze unstructured text using AI and Amazon Comprehend
* Create a chatbot and interact with it using speech and text input
* Retrieve external data via your chatbot
* Develop a natural language interface
* Apply AI to images and videos with Amazon Rekognition

**Audience & Pre-Requisites**

This course is geared for attendees with basic Linux and ML skills who wish to use artificial intelligence and machine learning on AWS to create engaging applications

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Introduction to Amazon Web Services**

* Introduction
* What is AWS?
* What is Amazon S3?
* Core S3 Concepts
* Data Replication
* AWS Command-Line Interface (CLI)
* Command Line-Interface (CLI) Usage
* Recursion and Parameters
* Using the AWS Console to Identify Machine Learning Services

1. **Summarizing Text Documents Using NLP**

* Introduction
* What is Natural Language Processing?
* Using Amazon Comprehend to Inspect Text and Determine the Primary Language
* Extracting Information in a Set of Documents
* Setting up a Lambda function and Analyzing Imported Text Using Comprehend

1. **Perform Topic Modeling and Theme Extraction**

* Introduction
* Extracting and Analyzing Common Themes

1. **Creating a Chatbot with Natural Language**

* Introduction
* What is a Chatbot?
* What is Natural Language Understanding?
* Setting Up with Amazon Lex
* Creating a Custom Chatbot
* A Bot Recognizing an Intent and Filling a Slot
* Lambda Function – Implementation of Business Logic

1. **Using Speech with the Chatbot**

* Introduction
* Amazon Connect Basics
* Interacting with the Chatbot
* Talking to Your Chatbot through a Call Center using Amazon Connect
* Using Amazon Lex Chatbots with Amazon Connect

1. **Analyzing Images with Computer Vision**

* Introduction
* Amazon Rekognition Basics
* Rekognition and Deep Learning

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.



CATEGORY CyberSecurity

# Machine Learning for Cybersecurity

**Course Snapshot**

* **Course:** Machine Learning for Cybersecurity
* **Duration:** 3 days
* **Skill-level**: Foundation-level Machine Learning for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to get into the world of smart data security using machine learning algorithms and Python libraries
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

Cyber threats today are one of the costliest losses that an organization can face. In this course, we use the most efficient tool to solve the big problems that exist in the cybersecurity domain. The course begins by giving you the basics of ML in cybersecurity using Python and its libraries. You will explore various ML domains (such as time series analysis and ensemble modeling) to get your foundations right. You will implement various examples such as building system to identify malicious URLs and building a program to detect fraudulent emails and spam. Later, you will learn how to make effective use of K-means algorithm to develop a solution to detect and alert you to any malicious activity in the network. Also learn how to implement biometrics and fingerprint to validate whether the user is a legitimate user or not. Finally, you will see how we change the game with TensorFlow and learn how deep learning is effective for creating models and training systems

Working in a hands-on learning environment, led by our Machine Learning with Azure expert instructor, students will learn about and explore:

* Learn machine learning algorithms and cybersecurity fundamentals
* Automate your daily workflow by applying use cases to many facets of security
* Implement smart machine learning solutions to detect various cybersecurity problems

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Use machine learning algorithms with complex datasets to implement cybersecurity concepts
* Implement machine learning algorithms such as clustering, k-means, and Naive Bayes to solve real-world problems
* Learn to speed up a system using Python libraries with NumPy, Scikit-learn, and CUDA
* Understand how to combat malware, detect spam, and fight financial fraud to mitigate cyber crimes
* Use TensorFlow in the cybersecurity domain and implement real-world examples
* Learn how machine learning and Python can be used in complex cyber issues

**Audience & Pre-Requisites**

This course is geared for attendees wants to get into the world of smart data security using machine learning algorithms and Python libraries

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, Microsoft azure and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Basics of Machine Learning in Cybersecurity**

* What is machine learning?

1. **Time Series Analysis and Ensemble Modeling**

* What is a time series?
* Classes of time series models
* Time series decomposition
* Use cases for time series
* Time series analysis in cybersecurity
* Time series trends and seasonal spikes
* Predicting DDoS attacks
* Ensemble learning methods
* Voting ensemble method to detect cyber attacks

1. **Segregating Legitimate and Lousy URLs**

* Introduction to the types of abnormalities in URLs
* Using heuristics to detect malicious pages
* Using machine learning to detect malicious URLs
* Logistic regression to detect malicious URLs
* SVM to detect malicious URLs
* Multiclass classification for URL classification

1. **Knocking Down CAPTCHAs**

* Characteristics of CAPTCHA
* Using artificial intelligence to crack CAPTCHA

1. **Using Data Science to Catch Email Fraud and Spam**

* Email spoofing
* Spam detection

1. **Efficient Network Anomaly Detection Using k-means**

* Stages of a network attack
* Dealing with lateral movement in networks
* Using Windows event logs to detect network anomalies
* Ingesting active directory data
* Data parsing
* Modeling
* Detecting anomalies in a network with k-means

1. **Decision Tree and Context-Based Malicious Event Detection**

* Adware
* Bots
* Bugs
* Ransomware
* Rootkit
* Spyware
* Trojan horses
* Viruses
* Worms
* Malicious data injection within databases
* Malicious injections in wireless sensors
* Use case
* Revisiting malicious URL detection with decision trees

1. **Catching Impersonators and Hackers Red Handed**

* Understanding impersonation
* Different types of impersonation fraud
* Levenshtein distance

1. **Changing the Game with TensorFlow**

* Introduction to TensorFlow
* Installation of TensorFlow
* TensorFlow for Windows users
* Hello world in TensorFlow
* Importing the MNIST dataset
* Computation graphs
* Tensor processing unit
* Using TensorFlow for intrusion detection

1. **Financial Fraud and How Deep Learning Can Mitigate It**

* Financial Fraud and How Deep Learning Can Mitigate It
* Machine learning to detect financial fraud
* Logistic regression classifier – under-sampled data
* Deep learning time

1. **Case Studies**

* Introduction to our password dataset

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.

# Applied Dark Web Analysis

**Course Snapshot**

* **Course:** Applied Dark Web Analysis
* **Duration:** 3 days
* **Skill-level**: Foundation-level Dark Web for Intermediate skilled team members. This is not a basic class.
* **Targeted Audience**: This course is geared for those who wants to Understand the concept Dark Web and Dark Net to utilize it for effective cybersecurity
* **Hands-on Learning:** This course is approximately 50% hands-on lab to 50% lecture ratio, combining engaging lecture, demos, group activities and discussions with machine-based student labs and exercises. Student machines are required.
* **Delivery Format:** This course is available for onsite private classroom presentation.
* **Customizable:** This course may be tailored to target your specific training skills objectives, tools of choice and learning goals.

The overall world wide web is divided into three main areas - the Surface Web, the Deep Web, and the Dark Web. The Deep Web and Dark Web are the two areas which are not accessible through standard search engines or browsers. It becomes extremely important for security professionals to have control over these areas to analyze the security of your organization. This course will initially introduce you to the concept of the Deep Web and the Dark Web and their significance in the security sector. Then we will deep dive into installing operating systems and Tor Browser for privacy, security and anonymity while accessing them. During the course of the course, we will also share some best practices which will be useful in using the tools for best effect. By the end of this course, you will have hands-on experience working with the Deep Web and the Dark Web for security analysis

Working in a hands-on learning environment, led by our Dark Web expert instructor, students will learn about and explore:

* Understand the concept of Dark Net and Deep Web
* Use Tor to extract data and maintain anonymity
* Develop a security framework using Deep web evidences

**Topics Covered**: This is a high-level list of topics covered in this course. Please see the detailed Agenda below

* Access the Deep Web and the Dark Web
* Learn to search and find information in the Dark Web
* Protect yourself while browsing the Dark Web
* Understand what the Deep Web and Dark Web are
* Learn what information you can gather, and how

**Audience & Pre-Requisites**

This course is geared for attendees wants to Understand the concept Dark Web and Dark Net to utilize it for effective cybersecurity

**Pre-Requisites:** Students should have

* Basic to Intermediate IT Skills, Microsoft azure and Machine Learning knowledge
* Good foundational mathematics or logic skills
* Basic Linux skills, including familiarity with command-line options such as ls, cd, cp, and su

**Course Agenda / Topics**

1. **Understanding the Deep and Dark Web**

* The origin of the internet
* The Deep Web
* The Dark Web

1. **Working with the Deep Web**

* Maintaining privacy on the Dark Web
* Transacting on the Dark Web
* Deep web emails - Onion Mail

1. **The Future of the Dark Web**

* What does the future of the Deep Web hold for us?
* Dark Web markets
* The TOR Project
* Public interest in the Dark Web

1. **Installing a Linux Virtual Machine (VM)**

* Linux Distributions
* Installing and using a Linux VM
* What else can we do with a Linux VM?

1. **Accessing the Dark Web with Tor Browser**

* What is Tor Browser?
* Installing Tor on Linux
* Tor Project recommendations on the safe use of Tor

1. **Installing Tails OS**

* What is Tails OS?
* Tails OS installation prerequisites
* Downloading Tails OS
* Installing Tails OS

1. **Installing Whonix**

* What is Whonix?
* Whonix installation and prerequisites
* Using Tor Browser with Whonix

1. **Installing Qubes OS**

* What is Qubes OS?
* Accessing the Dark Web with Qubes

1. **What Goes on in the Dark Web - Case Studies**

* The good and evil side of the Dark Web
* Onion websites
* Illegal sales on the Dark Web using Bitcoin

1. **The Dangers of the Dark Web**

* Online scams
* Avoiding the risks on a Dark Web market
* Dangers of the Dark Web

1. **Using the Dark Web for Your Business**

* IT professionals
* Business companies
* Law enforcement agencies
* Military organizations
* Cybersecurity professionals

**Student Materials:** Each student will receive a **Student Guide** with course notes, code samples, software tutorials, diagrams and related reference materials and links (as applicable). Our courses also include step by step hands-on lab instructions and and solutions, clearly illustrated for users to complete hands-on work in class, and to revisit to review or refresh skills at any time. Students will also receive the project files (or code, if applicable) and solutions required for the hands-on work.