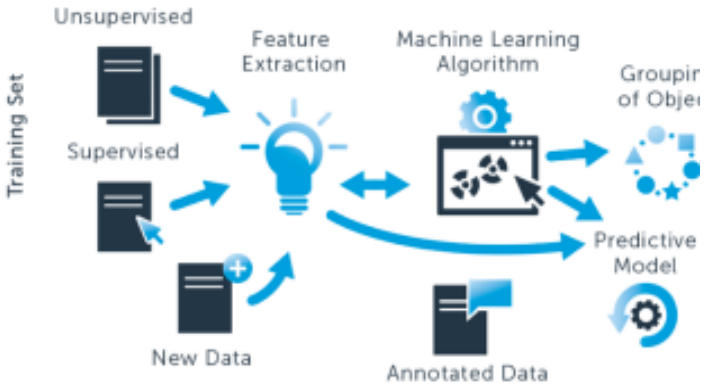

Machine Learning Artificial Intelligence

Applied Data Science

.

Course Content:

1. **Introduction to the Data Science**
2. **Python - Core, Advance, Analytics**
3. **Introduction to R Programming**
4. **Statistics and Probability – Business Analytics**
5. **Exploratory Data Analysis – Story telling using Visualization**
6. **Machine Learning – Statistics Decision making, Methods and Algorithms**
7. **Tableau For Story Telling**
8. **Natural Language Processing – Text Mining**
9. **Artificial Intelligence – Deep learning and Computer vision**
10. **Applying ML to Big Data using Hadoop and Spark**
11. **Deployment**
12. **Classroom Session and Lab Session**
13. **Capstone Projects: 4 - 5 on Retail, Banking, Insurance, Health, Image Processing and Text Mining.**

1. Introduction Data Science and Data

- Introduction to data science
- Application of Data Science
- Life cycle of data science projects
- Data and its various forms

2. Python – Core, Advance & Data Analysis

Module – 1: Basics of Python Data Structures

- Introduction to Python basic syntax
- Basics of Data types
  - Numbers, Variables, Strings
  - List, Tuple, Set, Dictionary

Module – 2: Python Statements, Methods, Functions and Expressions

- Python Statements
  - If Elif and Else Statements
  - For Loops, While Loops, Useful Operators
  - Methods and Functions in Python
  - Expression
  - Lambda Expression
  - Map, Filter Functions, Nested Statements and Scope

Module – 3: Advanced Python

- Modules and Packages
- File Handling, Errors and Exceptions Handing

Module – 4: More Advanced Python

- Object Oriented Programming, Inheritance, Polymorphism,
- Abstraction, Encapsulation, Regular Expression

Module - 5: Python for Data Analysis

- **Introduction of NumPy**
  - Arrays, Indexing, Array processing
  - Array Input and Output
- **Introduction to Pandas**
  - Series, DataFrames, Indexing and slicing, Groupby
  - Concatenating, Merging Joining, Missing Values
  - Operations, Data Input and Output
  - Pivot, Cross tab

3. Introduction to R Programming

Module - 1: Basics of R

- Data types, Variables, Operators
- Strings, List
- Decision making
- Loops, Vectors
- Arrays, Functions

Module – 2: R for Data Analysis

- Importing Data from texts and Spreadsheets
- Data Frames
- Packages, Libraries
- Data Manipulation and reshaping
- Data Visualization using R
- Data Transformation

4. Statistics and Probability

Module -1: Descriptive Statistics

- Understanding the properties of attribute
- Central tendencies (Mean, Median, Mode)
- Measure of Spread (Range Variance Standard Deviation)
- Basics of Probability
- Expectation and Variance of a variable
- Z- test
- Probability theory
- Random Variables
- Probability theory
- Conditional Probability
- Bayes theorem
- Deeper into probability distribution
- Discrete Probability Distribution:
- Bernoulli, Binomial, Geometric, Poison and properties of each.
- Continuous Probability Distribution:
    - Exponential
    - Normal distribution and t- distribution

Module -2: Inferential Statistics

- **Judgments and Conclusion from samples.**
    - Inferential Statistics:
        - Population from a sample and vice versa

- Central Limit Theorem
- Sampling Distribution
- Confidence Interval, Hypothesis Testing.

▪ More Statistical testing:

- chi-square test,
- t-test, ANOVA

5. Exploratory Data Analysis using Python

Module -1:

- **Data Preprocessing using Python with usecase.**
  - Data-point, vector, observation
  - Dataset
  - Input variables/features/dimensions/independent variable ○ Output Variable/Class Label/ Response Label/ dependent variable ○ Objective: Classification.
  - Standardization and Normalization
  - Label encoding (Type conversion)
  - Splitting data

Module -2:

- **EDA with Visualization using Seaborn and Matplotlib**
  - Scatter-plot: 2D, 3D.
  - Pair plots.
  - PDF, CDF, Univariate analysis.
  - Histogram and PDF
  - Univariate analysis using PDFs.
  - Cumulative distribution function (CDF)
  - Mean, Variance, Std-dev
  - Median, Percentiles, Quantiles, IQR, MAD and Outliers,
  - Box-plot with whiskers,
  - Violin plots.
  - Summarizing plots.
  - Univariate, Bivariate and Multivariate analysis.
  - Multivariate probability density, contour plot.

6. Machine Learning – Statistical Decision, Algorithms

Module-1: Predictive Analysis

- **Linear Regression**
  - Relationship between variables: Regression (Linear, Multivariate Linear Regression) in prediction.
  - Understanding the summary output of Linear Regression
  - Residual Analysis
  - Identifying significant features, feature reduction using AIC, multi-collinearity check,

observing influential points, etc.
  - Hypothesis testing of Regression Model
  - Confidence intervals of Slope
  - R-square and goodness of fit
  - Influential Observation – Leverage
  - Multiple Linear Regression and Polynomial Regression
  - Categorical Variable in Regression
- **Logistic Regression** (Logit function) and interpretation
  - Hands-on Python Session on Logistic Regression using business case.
  - ROC
- **Naïve Bayes classifier**
  - Review probability distributions, Joint and conditional probabilities
  - Model Assumptions, Probability estimation
  - Required data processing
  - Feature Selection
  - Classifier
- **Time Series Analysis:**
  - Trend analysis
  - Cyclical and Seasonal analysis
  - Smoothing; Moving averages; Auto-correlation; **ARIMA**
  - Application of Time Series in financial marker

Module-2: Algorithms in Machine Learning

- **Rule based in Supervised Learning**
  - **Classification Rules**
  - **Decision Tree –** Indirect
  - Direct: Sequential covering
  - Decision nodes and leaf nodes
  - Variable Selection, Parent and child nodes branching
  - Stopping Criterion, Tree pruning, Depth of tree, Over fitting
  - Metrics for decision trees-Gini impurity, Information Gain, Variance Reduction
  - Regression using decision tree
  - Interpretation of a decision tree using If-else
  - Accuracy estimation using cross-validation

- **Distance Based Approach (k-NN)**
  - What is KNN and why do we use it?
  - KNN-algorithm and regression
  - Curse of dimensionality and brief introduction to dimension reduction
  - KNN-outlier treatment and anomaly detection
  - Cross-Validation
  - Pros and cons of KNN
- **Mathematical Approach - Support Vector Machine (SVM)**
  - Linear learning machines and Kernel space, making kernels and working in feature space

- **Ensemble Models**
  - Introduction to Ensemble
  - Bias and Tradeoff
  - Bagging & boosting and its impact on bias and variance
  - Random forest
  - Gradient Boosting Machines and **XGBoost**

Module-3: Unsupervised Learning

- **Clustering**
  - Different clustering methods
  - review of several distance measures
  - Iterative distance-based clustering
  - Dealing with continuous, categorical values in **K-Means**
  - Constructing a **hierarchical cluster**, and **density-based** clustering.
  - Test for stability check of clusters

- **Recommendation System**
  - Association rules
    - How to combine clustering and classification;
    - A mathematical model for association analysis
    - Apriori: Constructs large item sets with mini sup by iterations
    - Metrics of rules-Lift, Support, Confidence, Conviction
  - Collaborative filtering and its applications areas

7. Tableau for Reporting

Module – 1: Tableau Basics

- The Business Challenge
- Connecting Tableau to a Data File
- Navigating Tableau, Creating Calculated Fields
- Adding Color, Adding Labels and Formatting, Exporting Your Worksheet

Module – 2: Time Series, Aggregating, and Filters

- Working with Data Extracts in Tableau
- Working with Time Series
- Understanding, Aggregating, Granularity and Level and Detail
- Creating an Area Chart & Learning About Highlighting
- Adding a Filter and Quick Filter

Module – 3: Dashboard

- Joining Data in Tableau
- Creating a Map, working with Hierarchies
- Creating a Scatter Plot, Applying Filters to Multiple Worksheets
- Create First Dashboard
- Adding an Interactive Action – Filter
- Adding an Interactive Action – Highlighting

Module – 4: Calculations, Advanced Dashboards, Storytelling

- Downloading the Dataset and connecting to Tableau
- Creating Table Calculations
- Creating Bins and distributions
- Leveraging the Power of Parameter
- How to Create a Tree Map Chart
- Creating a Customer Segmentation Dashboard
- Advance Dashboard Interactivity
- Creating a Storyline

8. Natural Language Processing (NLP) – Text Mining

- Introduction to Text Mining and its Application
  - Introduction to NLTK, Spacy libraries in python
- Structured and Unstructured Data
- Extracting Unstructured text from files and Websites
- Processing with Raw Text
  - Regular Expression for Detecting Word Patterns
  - Normalizing Text
  - Tokenizing Text
  - Segmentation
    - Stemming and Lemma
- Categorizing and Tagging
  - Automatic Tagging
  - N-Gram Tagging
  - Transformation based tagging

- Introduction to the Fundaments of information retrieval
  - TF and IDF
  - Bag-of-words
  - Thinking about the math behind text; Properties of words; Vector Space Model ○ Named Entity Recognition
  - Relation Extraction

- Matrix factorization: Singular Value Decomposition (SVD)
  Text Indexing
  - Inverted Indexes
  - Boolean query processing
  - Handling phrase queries, proximity queries

- Latent Semantic Analysis
- Text classification
- Sentiment analysis

9. Artificial Intelligence – Deep Learning and Computer Vision

Deep Learning:

Module – 1: Introduction to Neural Networks

- Introduction to Neural Network
- Introduction to Perceptron
- Activation Functions
- Cost Functions
- Gradient Decent
- Stochastic Gradient Descent
- Back propagation

Module – 2: Tensorflow

- Tensorflow Basic Syntax
- Tensorflow Graphs
- Variables and Placeholder
- Saving and Restoring Models
- Tensorboard

Module – 3: Building Neural Network With TensorFlow

- Neural Network for Regression
- Neural Network for Classification
- Evaluating the ANN
- Improving and tuning the ANN

Module - 4: Convolutional Neural Networks (CNN)

- Convolution Operation
- ReLU Layer
- Pooling
- Flattening
- Full Connection
- Softmax and Cross Entropy

Module -5: Building Convolution Neural Network in Python

Computer Vision:

- OpenCV library in Python
- Getting Started with Images/Videos

- Operations on Images
- Image Processing in OpenCV
- Geometric Transformation of Images
- Rotation
- Affine Transformation
- Perspective Transformation
- Image Thresholding
- Contours
- Edge detections
- Morphological Transformation
- Harris Corner Detection
- Reshaping Images
- Normalizing Images
- Building Convolutional Network with Tensorflow
- Training CNN for Image Classification

Module -6: Keras (Backend Tensorflow)

- **Keras vs Tensorflow**
- Introduction to Keras
- Building Artificial Neural Network with Keras
- Building Convolution Neural Network with Keras

Module -7: Recurrent Neural Networks (RNN)

- The Idea behind Recurrent Neural Networks
- Vanishing Gradient Problem
- LSTM (Long Short Term Memory)

Module -8: Building Recurrent Neural Networks with Tensorflow and Keras

- Time Series
- Time Series Forecasting with LSTM

10. ML to Big Data using Hadoop and Spark Ecosystem

- Structured vs unstructured data
- 4 Vs if Big Data
- Hadoop Ecosystem
- Applications of Big Data
- Introduction to Map-Reduce
- Hadoop Ecosystem (Pig, Hive etc.)
- Introduction to Spark and Scala
- Spark ML
- Spark SQL

11. Deployment

At the end of any Data Science project end with Deploying.

- Creating pickle and frozen files
- Cloud Deploying Machine Learning and Deep Learning model for production