**Implement Machine Learning for Data Science**

**Duration: 5 days (40 Hrs)**

Data analytics is often used by companies to search for trends in their growth. It often uses data insights to make an impact by connecting the dots between trends and patterns while data science is more about just insights.

Machine learning and data analytics are a part of data science. Because the machine learning algorithm obviously depends on having data to learn, data science is a broader term and does not only focus on implementing algorithms and statistics but also includes the entire data processing methodology. Thus, data science is a broader term that could incorporate multiple concepts like data analytics, machine learning, predictive analytics, and business analytics.

**Course Overview:**

Get Introduced to Machine Learning with Supervised, Unsupervised, Reinforcement Learning

Get idea about how to use Natural Language Processing, Artificial Neural Networks

Understand the use of NumPy, pandas and sckit packages.

Perform exploration of data and understand distribution of data, summary statistics, inferential statistics, different types of sampling, correlation.

**Pre-requisite:**

Basic Python Syntaxes

Comfortable using Jupyter notebooks.

Loops and Conditional Statements

Writing Functions and using lambda expressions in Python

**Target Audience:**

Audience interested in Machine Learning.

Any audience who are comfortable with coding but who are interested in Machine Learning and want to apply it easily on datasets.

Any audience who wants to create added value to their business by using powerful ML tools.

**Session Objectives Program Structure**

**Introduction To Data Science**

* What is Data Science?
* Data Science team structure
* Role of a data scientist
* Introduction to data science discipline
* Data scientist responsibilities, skills and knowledge Presentations, discussions

**Introduction to Machine Learning**

* Patterns in data, what does it mean?
* Representing reality in models
* What is Machine Learning?
* Requisites for Machine Learning

**Introduction to various forms of patterns in data such as dependency, associations, clusters, trends, classes etc.**

* What is a model? How it represents the real-world Machine learning, a blend of coding and mathematics
* Presentations, discussions
* Preparing for Machine Learning projects
* Defining the objectives
* Identifying the required data items
* Identifying sources of data
* Data cleansing
* Preparing data for ML

**Introduce the approach to machine learning project with focus on clarity of objectives, identifying sources of data, preparing data for analytics Hadoop stack and its applications**

* Presentations, discussions
* Machine Learning Styles
* Supervised machine learning
* Unsupervised machine learning
* To familiarize participants with the two broad classification of machine learning styles
* Presentations and discussion on live case studies
* Python – a quick overview
* Installing Python, Anaconda
* Python packages for data science
* Python data structures (lists, dictionaries, arrays and Dataframes)

**Introduce participants to the basics of Python. Participants will be familiarized with Python data structures and how to use them for data transformations.**

* Presentations, discussions
* Data Interfaces
* CSV files
* Excel files
* Text files
* Databases
* Web links

**Introduce participants to the Python way of loading multiple types of data from various sources.**

* Presentations, discussions
* Descriptive / Basic Statistics using Python.
* Data about data, studying individual attributes of data.
* Univariate analysis using Mean, Median, Mode, Variance
* Interpreting basic statistics reports

**Inferential Statistics**

* A thorough introduction to metrics about data such as mean, median, quantiles, mode, relation between variables
* Familiarize participants to Python functions for basic statistical analysis.
* Presentations, discussions
* Normal distribution and it’s characteristics.
* P Value

**Sample, sampling and population.**

* How sample metrics are used to infer about population parameters
* What is a hypothesis, Null and Alternate hypothesis?
* Presentations, discussions
* This will not be an in depth mathematical session.
* Hypothesis and hypothesis testing
* Infer about population from sample.
* Accepting and rejecting alternate hypothesis

**Type I and Type II errors.**

* Feature Engineering
* Multivariate analysis, interpreting
* R & R Square analysis
* Selecting attributes for dimensions, independent dimensions
* Dimensionality reduction using high correlation filters, low variance filters, Mutual information filters, Principal component analysis the challenge of selecting the most appropriate set of dimensions to build a model.
* Curse of dimensionality, its meaning and how to address the problem.

**Supervised Learning Methods**

* **Regression**
* Linear regression
* Multiple Linear Regression
* Polynomial Regression
* Decision trees
* Random Forest Regressor
* **Classification**
* Logistic Regression
* K-Nearest Neighbours
* Support Vector Machine
* Kernel-Support Vector Machine
* Decision Tree Classifier
* Random Forest Classifier
* Naive Bayesian classifiers
* Introduce participants to supervised learning approach with focus on model generation through training data, testing the model, interpreting the results.
* Presentations, discussions, and hands-on coding in Python

**Unsupervised Learning Methods**

* K Means Clustering
* To explain concepts of unsupervised machine learning, their applications and how systems learn on their own.
* Presentations, discussions, and hands-on coding in Python

**Reinforcement Learning Learning Methods**

* Thompson Sampling
* Upper Confidence Bound

**Introduction to Deep Learning**

* Implementing and understanding Artificial Neural Network
* Brief introduction to Neural Network using Karas.

**Modelling errors and resolutions**

* Bias and variance errors
* Overfit and underfitting models
* Error functions and Stochastic
* Gradient Descent
* Kernels based methods for improving model performance.
* Concepts of over fitting / under fitting and generalization.
* Bias and variance errors and dealing with these errors to improve model accuracy.
* Error functions, problem of local minima and gradient descent
* Need for kernel-based methods in modelling and how they work.
* Presentations, discussions, concept only

Student projects Participants can select project from their environment or take up sample project from available list Participants work in groups to decide the most suitable machine learning algorithms, clean and transform data if required, build the models and tune the same.