

DATA SCIENCE COURSE CONTENT

Faculty:

Fee:

Duration:

Demo Time:

Batch Time:

INTRODUCTION:

- ❖ What is Data Science?
- ❖ What is Machine Learning?
- ❖ Who are Data Scientists?
- ❖ What skills are required to become a Data Scientist?
- ❖ Why should a data scientist learn both R as well As Python?
- ❖ Is there a importance for Statistics in Data Science ?
- ❖ Where we implement Data Science in real time?
- ❖ In which sectors of business we implement Data Science?
- ❖ What kind of Projects they work on?
- ❖ The glimpse of the technical and the managerial functions of a Data Science
- ❖ Exposure to marketing scenario and customer dynamics
- ❖ Acquiring knowledge about the random processes involved in trading
- ❖ Data quality
- ❖ Data security concepts
- ❖ Each and Every technique will be explained with a Real time Example.

INTRODUCTIONS TO MACHINE LEARNING TECHNIQUES:

❖ Supervised Learning

- What is Supervised learning?
- What Algorithms comes under Supervised learning?

❖ Simple Linear Regression

❖ Simple Linear Regression with Gradient Descent

❖ Linear Regression with Quadratic equation

❖ Stochastic Gradient Descent Model

❖ Error Minimization Model

❖ Linear Regression for classification

- Estimating the coefficients
- Assessing the accuracy of the coefficient estimates
- Multiple Linear regression
- Finding a new dependent Variable as prediction
- How to find a Slope and Intercept

❖ Logistic Regression

- Spline Regression
- Sigmoid, Rigid and Lasso Regression
- The Logistic Model
- Estimating the Regression Coefficients
- Making predictions
- Multiple Logistic Regression

❖ Non-Linear Regression

❖ Linear Discriminant Analysis

- Exploratory Data Analysis
- Single Variant Model Analysis

- Multi Variant Model Analysis
- Quadratic Discriminant Analysis
- ❖ Naïve bayes
- ❖ Support Vector Machines
 - Maximum Margin Classifier
 - What is a hyperplane
 - What are marginal Hyper planes
 - Classification using a separating Hyper plane
- ❖ SVM Linear Model
- ❖ SVM Non-Linear Model
- ❖ SVM Kernel Method
- ❖ SVM Power Kernel Method
- ❖ K-Nearest Neighbors
 - Advantages of KNN
 - Disadvantages of KNN
 - How to find an appropriate K value?
 - How to we use KNN for classification?
 - How do we use KNN for regression?
 - How to find the Nearest Neighbor
- ❖ Ada-Boost / Adaptive- Boosting
- ❖ Random Forest
 - What is Random Forests?
 - Overview of Random Forests
 - Classification with Random Forests
 - Implementation of Random Forests
- ❖ Decision trees
 - How to build a decision tree
 - What is classification and its uses?

- What is a decision tree
 - Creating a Decision tree
 - Information Entropy
 - Information Gain
 - Examples
- ❖ Gini Index and Entropy methods
 - ❖ Extra Trees
 - ❖ Gradient Boosting Method
 - ❖ Trees for prediction
 - ❖ Trees for classification
 - ❖ Taxonomy Creation
 - ❖ Sentiment analysis
 - ❖ Text classification models
 - ❖ Process of tree building
 - ❖ Problem of Underfitting
 - ❖ Problem of overfitting
 - ❖ Pruning a tree back
 - ❖ Lasso & Rigid Regression
 - ❖ Unsupervised Learning
 - ❖ K-means Clustering
 - ❖ Clustering
 - ❖ Clustering techniques
 - ❖ Hierarchical clustering
 - ❖ Time Series
 - Qualitative Methods
 - Quantitative Methods
 - ❖ Moving Averages
 - ❖ Smoothing Techniques
 - ❖ Exponential Smoothing

- ❖ **Weighted Moving Averages**
- ❖ **ARMA**
- ❖ **ARIMA**
- ❖ **Trend and Seasonality forecasting**
 - **Linear**
 - **Quadratic**
 - **Exponential**
 - **Auto Regressive**
- ❖ **Semi Supervised Learning / Reinforcement Learning**

- ❖ **Market basket Analysis**
 - **Characteristics of Market Basket Analysis**
 - **Applications of Market basket Analysis**
 - **Association rules**
- ❖ **Forecasting Analysis**
 - **Lacuna in statistical forecasting**
 - **Why Forecasting?**
 - **Features common to all Forecasts**
 - **Elements of a good Forecast**
 - **Types of Forecast**
 - **Qualitative Methods**
 - **Quantitative Methods**
 - **Techniques for Averaging**
 - **Techniques for Trend**
 - **Techniques for Seasonality**
 - **Seasonalizing Trend**
 - **De-seasonalizing Trend**
- ❖ **Predictive Analysis**
- ❖ **Various approaches to solve a problem**

- ❖ Pros and Cons of Machine Learning Algorithms
- ❖ Data Leakage in Machine Learning
 - Re- Sampling and Ensembling Techniques
- ❖ Bagging
- ❖ Boosting
- ❖ RF techniques

Principal Component Analysis

- ❖ Advantages of Principal Components
- ❖ Applications of PCA

BASIC STATISTICS & BUSINESS STATISTICS:

- ❖ Basic Statistics and Advanced Statistics
- ❖ Mean, Median, mode
- ❖ Standard Deviation
- ❖ Probabilities
- ❖ Permutations and Combinations
- ❖ Distributions
- ❖ Correlations
- ❖ Covariance
- ❖ Distance formulas
- ❖ Similarities
- ❖ Minkowski Distance
- ❖ L1 Norm
- ❖ Lmin Norm
- ❖ Manhattan Distance
- ❖ Taxi cab

- ❖ **Supremum**
- ❖ **Lmax Norm**
- ❖ **Euclidean Distance**
- ❖ **Cosine Similarity**
- ❖ **Hamming Distance**
- ❖ **Leveinsthein Distance**
- ❖ **Simple Matching Coefficients**
- ❖ **Jaccards Equation**
- ❖ **Differences between Distances and Similarities**
- ❖ **Product of Dot Vectors**
- ❖ **Binary values**
- ❖ **Mahalnobis Distance**
- ❖ **Matrix Multiplication**
- ❖ **Variance**
- ❖ **Conditional Probabilities**
- ❖ **Difference between Statistics and probability**
- ❖ **Confusion Matrix**
- ❖ **Error Types**
- ❖ **Mean Squared Error**
- ❖ **Root Mean Square Error**
- ❖ **Accuracy Measures**

DATA TYPES

- ❖ **Continuous Variables**
- ❖ **Ordinal Variables**
- ❖ **Categorical Variables**
- ❖ **Time Series**
- ❖ **Miscellaneous**
- ❖ **Descriptive Statistics**
- ❖ **Sampling**
- ❖ **Random Sampling**

- ❖ Systematic Sampling
- ❖ Stratified Sampling

Data Distributions

- ❖ Normal Distribution
- ❖ Binomial Distribution
- ❖ Weibull Distribution
- ❖ Poison Distribution
- ❖ Exponential Distribution

Inferential statistics

- ❖ Error types
- ❖ Null and alternate hypothesis
- ❖ Reject and acceptance criterion

INTRODUCTION TO PYTHON:

- Installation of python
- ❖ Variables and Types
- ❖ Data Structures
- ❖ Lists
- ❖ Dictionaries
- ❖ Tuples
- ❖ Strings
- ❖ Conditional functions

Working with Python Libraries for Machine Learning:

- ❖ Installing Libraries
- ❖ Pandas
- ❖ Numpy
- ❖ SKlearn

- ❖ **Mathplotlib**
- ❖ **Scipy**
- ❖ **How to create a Data Frame**
- ❖ **Grouping on the Class variable**
- ❖ **Converting different formats**
- ❖ **Working on all Machine Learning Techniques in Python.**
- ❖ **Machine Learning terminology for model Building.**
- ❖ **Feature Selections and Preprocessing**
- ❖ **How to select the right data**
- ❖ **Which are the best features to use**
- ❖ **Additional feature selection techniques**
- ❖ **Preprocessing Introduction**
- ❖ **Preprocessing Scaling Techniques**
- ❖ **How to process data**
- ❖ **How to scale data**
- ❖ **Standardizing the scalar**
- ❖ **Pipeline techniques for scaling and standardizing**
- ❖ **Which Algorithm perform Best**
- ❖ **Highly efficient ML Algorithms**
- ❖ **Bagging Decision Trees**
- ❖ **Ensembling Techniques**
- ❖ **Boosting Techniques**
- ❖ **How to consider which algorithm is best**
- ❖ **Model Selection**
- ❖ **Cross validation**
- ❖ **Model tuning**
- ❖ **Parameter tuning with GridSearchCV**
- ❖ **Algorithm Tuning**
- ❖ **How to compare ML algorithms**

INTRODUCTION TO TABLEAU:

- ❖ Features of Tableau
- ❖ Connecting Tableau to a Data File
- ❖ Navigating Tableau – Dimensions and Measure
- ❖ Creating a Calculating Field
- ❖ Adding Colors
- ❖ Adding Labels
- ❖ Formatting
- ❖ Creating groups and Bins
- ❖ Combining two measures in a single chart
- ❖ Working with Hierarchies
- ❖ How to do Sorting
- ❖ Dual Axis Chart
- ❖ Doing Analytics in Tableau
- ❖ How to forecast using tableau
- ❖ How to Model a Trend Line
- ❖ How to custom Distribution band
- ❖ How to custom a reference line
- ❖ How to summarize Totals, Constant line and Average lines
- ❖ Calculations in Tableau
- ❖ Sort and Filters
- ❖ Charts
- ❖ Bar Graphs
- ❖ Pie Charts
- ❖ Line chart
- ❖ Bubble Chart
- ❖ Scatter Plot
- ❖ Box plot
- ❖ Tree Map

- ❖ Histogram
- ❖ Waterfall chart
- ❖ Gantt chart
- ❖ Cross Tab
- ❖ Tableau Dashboard

INTRODUCTION TO R:

- Installing R
- Installing Rstudio
- Installing Libraries
- ❖ Caret
- ❖ Rattle
- ❖ MLbench
- ❖ Mass
- ❖ e1071
- ❖ Corrplot etc.,

Data Structures

Creating Variables

- ❖ Lists
- ❖ Matrices
- ❖ Arrays
- ❖ If Conditions

MODEL VALIDATION AND DEPLOYMENT:

- ❖ Error measurement
- ❖ RMSE, MSE
- ❖ Misclassification Rate
- ❖ Confusion Matrix

- ❖ AUC (Area Under the Curve)
- ❖ ROC curve
- ❖ How to create a Data Frame
- ❖ Grouping on the Class variable
- ❖ Converting different formats
- ❖ Working on all Machine Learning Techniques in R.
- ❖ Machine Learning terminology for model Building.
- ❖ Which Algorithm perform Best
- ❖ Highly efficient ML Algorithms
- ❖ Bagging Decision Trees
- ❖ Ensembling Techniques
- ❖ Boosting Techniques
- ❖ How to consider which algorithm is best
- ❖ Re-Samplings
 - Cross Validation
 - Repeated Cross validation
 - Boot Strap
 - Leave One Out Cross Validation
- ❖ PreProcessing Transformations
 - Scale
 - Centre
 - Boxcox
 - YeoJohnson
 - Expotrans
 - Zv
 - NZV
 - PCA
 - ICA
 - Range
 - SpacialSign
 - ExpoTrans

❖ **Evaluation Metrics**

- **LogLoss**
- **Neg LogLoss**
- **RMSE**
- **MSE**
- **Kappa**

EXPLORATORY DATA ANALYSIS AND VISUALIZATION

- ❖ **Getting the data from the files to R and python Frames and connecting to Database.**
- ❖ **Cleaning and preparing the Data**
- ❖ **Handling missing values doing data manipulation**
- ❖ **Visualizing in R and Python (Plots and Charts)**
- ❖ **Adding more dimensions to the plots**
- ❖ **Visualizing in tableau**
- ❖ **Correlation**

DATA PROFILING:

- ❖ **Data profiling concepts**
- ❖ **Elicitation of data**
- ❖ **Data patterns (Trend, seasonality, cyclic effects etc.)**
- ❖ **Data analogies**
- ❖ **Data types (Deterministic, probabilistic, random)**
- ❖ **Dive deep into data quality and sanity**
- ❖ **Concepts of structured and unstructured data**
- ❖ **Real time data, temporal data and dynamic data**
- ❖ **Data governance**
- ❖ **Storage and retrieval**
- ❖ **Infrastructure**

TOOLS

- ❖ Basic concepts regarding the open source software – “R” and “Python”
- ❖ Simple illustrative examples to be tried to gain the familiarity with the
 - software and complete process in “Excel”

MACHINE LEARNING: (Both SYMANTICS and SYNTAX)

- ❖ We Teach all the algorithms not only in Python and R but we each all the back-end process in Ms-excel also.

