

Predictive Business Analytics

Duration 60 - 80 Hours

Topics covered under this program are:

1. Learning R:

- Introduction to Programming
- R and R- Studio
- Installation and Functioning of R and R-Studio
- Programming in R
- Functions in R
- R – Best Practices

2. Statistics and Analysis on R:

- Statistics and multiple data types
- Different Measurement Scales
- Measures of Variables and Central Tendency
- Sampling Theories
- Central Limit Theorem
- Normal Distribution
- Hypothesis Testing: Z/T/Chi Square Test

3. Linear Regression

- Introduction of Simple Linear Regression
- Multiple Linear Regression
- Building a Regression Model
 - Data Preparation
 - Model Building
- Assumptions for Linear Regression
- Interpretation of Outputs and Diagnostics
- Analyzing Model Fitment for your dataset
- Using Linear Model for predictions
- Model Improvements
 - Over fitting and Cross-Validation

- Multicollinearity and VIF
- Do it Yourself case studies
- Flavors of Advance Regression Models

4. Logistics Regression

- Linear Regression Vs. Logistics Regression
- Logistic Regression
 - Introduction to Classification and Challenges with Linear Regression
 - Event Rate and Class Bias
- Example Problem (Some real examples of Binary Classification problems), Mechanics and Mathematics behind Logistic Regression
- Assumptions for Logistic Regression
- Building a Logistic Regression Model
 - Data Preparation – Data Audit, Missing Value and Outliers
 - Variable Importance and Feature Extraction
 - Create WOE for Categorical Variables
 - Compute Information Value
- Multicollinearity (VIF)
- Building Logit Models
- Logistic Regression – Interpretation of Output
 - Coefficients
 - Variable Importance
- Model Diagnostics
 - Misclassification Error and Confusion Matrix
 - ROC Curve
 - Accuracy
 - Specificity, Sensitivity and F Score
 - Lift/Gain Charts and KS Curve
- Model Improvement
 - Over-fitting and Cross Validation
- Flavour of Advanced Classification Concepts – Classification of Unstructured Data
- Do it Yourself Case

5. Time Series Modeling

- Introduction to Time Series
 - Difference between Time Series, Cross-Sectional and Pooled Data

- Example Problem (Some real-world examples of Time Series Problems), Mechanics and Fundamental of Mathematics behind Time series Analysis
- Assumptions for Time Series analysis
- Understanding Time Series Data
 - Visualizing Time Series Data
 - Stationary vs. No Stationary Data
 - Trend vs Seasonality vs White Noise
- Decomposing Time Series Data
 - Decomposing Non-Seasonal Data
 - Decomposing Seasonal Data
 - Seasonally Adjusting
- Forecasts using Exponential Smoothing
 - Simple Exponential Smoothing
 - Holt's Exponential Smoothing
 - Holt-Winters Exponential Smoothing
 - Challenges with Smoothing
- ARIMA Models
 - Concept of Auto-Correlation and Partial Auto Correlation
 - Differencing a Time Series
 - Selecting a Candidate ARIMA Model
 - Forecasting Using an ARIMA Model
 - Predictions and Diagnostics
- Advanced Time Series Concepts
- Do it Yourself Case

6. Market Basket Analytics

- Supervised, Unsupervised and Semi-supervised Algorithms
- Concept of a Recommendation Engine
- Example Problem (Real world examples of MBA applications)
- MBA Hyper Parameters
 - Lift
 - Confidence
 - Support
- Generating output using Association rules
 - Filtration of Rules
 - Removal of Redundant Rules
 - Control the Rules

- Finding rules for Particular Entity
- Visualizing Rules
- Challenges with Association Rules and Ways to Overcome
- Advanced Recommendation Engine Concepts
- Do it Yourself Case

7. Decision Trees

- Type of Classification Algorithms
- Fundamentals of Tree based Systems
- Decision Boundary of Tree based Algorithms
- Types of Tree Algorithms
 - C4.5
 - CHAID
 - CART
- Concept of Impurity Measure
 - GINI
 - Chi Square
 - Entropy
- Building a Decision Tree Model
 - Data Preparation – Data Audit, Missing Value and Outliers
 - Creating Test and Training Samples
 - Variable Importance and Feature Extraction
- Prediction using Decision Trees
- Over fitting and Cross Validation
- Flavour of Advanced Concepts in Trees (Random Forests)

8. Clustering

- Type of Classification Algorithms
- Unsupervised Algorithms and Introduction to Clustering
 - Intro to Distance based Algorithms
- Example Problem (Some real-world examples of Clustering Applications)
- Assumptions for Clustering
- Mechanics of Clustering
 - Type of Measure- Euclidean, Manhattan, Jaccard
 - How are Clusters Generated
 - Using Clustering as a Base for Segmentation

- Creating Clusters
 - Standardization of Inputs
 - Deciding the Number of Clusters – Elbow Curve and Silhouette Distance
- Understanding the Output
 - Cluster Diagnosis
 - Profiling
- Advanced Clustering Concepts
- Do it Yourself Case

9. Python For Analytics

- Type of Classification Algorithms
- Understanding Python
 - Installing Python through Anaconda
 - Jupyter Notebook
 - Python Functions
 - Python Data types
 - Tuples
 - Lists
 - Dictionaries
 - Understanding the Data
 - Python Packages for Data Science
 - Importing and Exporting Data
 - Getting Started Analysing Data
 - The DataFrame Data Structure
 - DataFrame Indexing and Loading
 - Querying a DataFrame
 - Group by
 - Scales
 - Pivot Tables
 - Dealing with Missing Values
 - Data Formatting
 - Data Normalization
- Categorization in Python
 - Turning categorical variables into quantitative variables
 - Data Wrangling
 - Identifying Correlation

- Correlation - Visualization
 - Case Example - Classification
 - Exploratory Data Analysis
 - Using Logistic and Decision Tree for Model Development
 - Model Evaluation using Visualization
 - Model Evaluation
 - Overfitting, Underfitting and Model Selection

Contact Us: **TOLL FREE – 1800 200 5835**