

## Curriculum

### R Basics (40 Hours)

#### Introduction to R

10 Hours

R base software  
CRAN  
Rstudio the IDE  
Basic building blocks in R  
Sequence of numbers  
Vectors  
Matrices and data frames  
Logical statements  
Date and time functions

#### Basic Operations in R

2 Hours

Data, data types & variables  
Measures of central tendency in data  
Skewness in data  
Measures of dispersion  
Data distribution

#### Linear Regression

8 Hours

Covariance and correlation in data  
Multivariate analysis  
Assumptions of linearity  
Hypothesis testing  
Limitations of regression

#### Case Study For Linear Regression

Case for prediction problem

#### Logistic Regression

6 Hours

Reason for using logistic regression  
The logistic transform  
Logistic regression modelling  
Model optimisation  
Understanding the ROC curve

#### Case Study For Logistic Regression

Case for prediction problem

#### Decision Tree

4 Hours

Classification trees  
Regression trees

#### Case Study For Decision Tree

Case for prediction and classification

#### Segmentation

4 Hours

Clustering  
Kmeans algorithm  
Cluster size optimisation vs definition optimisation  
K- mediodid and fuzzy K means

#### Case Study For Segmentation

Case for clustering on bank customer data set

#### Association Rule Mining

2 Hours

Supervised vs unsupervised learning  
Recommendation engines  
Association rule mining  
Apriori algorithm  
Support, confidence and lift parameters

#### Case Study Market Basket Analytics

2 Hours

Case on hyper market shoppers dataset for MBA

### Advanced R (35 Hours)

#### Time Series

7 Hours

Time series decomposition  
Simple & Weighted moving average  
Single - Double exponential smoothing  
Triple exponential smoothing method  
AR, MA, ARMA, ARIMA, SARIMA models  
Using RMSE and MAPE for model performance

#### Case Study For Time Series

Case on automobile sales data

#### K Nearest Neighbours Algorithm

6 Hours

Lazy learning notion  
Computation of distance matrix  
The optimum K value  
Data transformations as a pre processing phase  
Model building on training data set  
Evaluation of model  
Advantages & disadvantages

#### Naïve Bayes Algorithm For Multi Class Predictions

6 Hours

Bayesian theorem  
Probabilities - the prior and posterior probabilities  
Conditional & Joint probabilities notion  
Naive approach - independence of features assumption  
Data processing - discretization of features  
Model building / testing / validation

#### Artificial Neural Networks

5 Hours

Understanding neural networks  
The biological inspiration  
The activation function  
The structure of network  
The ANN model  
Training the model  
Testing and validation

#### Support Vector Machines

5 Hours

Understanding SVM  
Concepts of linearly seperable vs non separable data  
Build, train, test and tune the model

#### Ensemble Models With Case Study

5 Hours

Understanding entropy  
Information value  
Model building on training data set  
Selecting the best split in data  
Pruning a decision tree  
Model validation on testing data set  
Improve model performance  
Bagging and boosting trees  
Random forests

### Job Readiness (15 Hours)

Interview Preparation - HR (4 Hours)

Interview Preparation - Domain (4 Hours)



Mock Interviews - HR and Domain (4 Hours)

Final Project Presentation (4 Hours)