



Machine Learning with R

(33 hours class room + 30 hours of practice sessions)

About the Course

The world is quietly being reshaped by machine learning. We no longer need to teach computers how to perform complex tasks like image recognition or text translation: instead, we build systems that let them learn how to do it themselves. R is very powerful open source software, which is the best tool for data analytics and machine learning, used by giant corporates including Google. In this course you will be learning how to use R and Machine Learning algorithms to solve business problems and extracting insights to enable to companies to stay one step ahead of their competitors.

Overview of the course

Class 1: Introduction to R Programming Language

Introduction and Installation of R software
R packages
Concepts of Vector – Numeric, Character, and Factor
Concepts of Data frame
Filtering
Usage of Boolean in Filtering
Sorting
Reshape of data using Tidyr package

Class 2: Data handling in R

Handling of Missing values
If else statement
Extra trick of using if else statement
Removal of Duplicates
Merging – Inner, Outer, Left and Right
Binding and Appending
Text functions
Data cleaning with efficient text functions
Inbuilt Numeric functions of R
Inbuilt String functions of R
Inbuilt other functions of R



Class 3: More data handling using R

Pivot Table of Excel in R

Table function

Count function of plyr package

Learning of SQL queries using R

Grouping numeric data

User defined functions (Macros) in R

Visualizing of Data

Class 4: Additional functions of R

Date functions with Lubridate package

Apply functions

User defined functions (Macros) in R

Visualizing of Data

Class 5: Statistics

Everything you want to know about statistics....Well sort of!!

Mean, Median, Mode

Standard Deviation, Variance,

Normal Distribution

Hypothesis testing

T-test, Anova, Normality test

Class 6: Linear Regression

Predictive Analytics – Linear Regression

Concepts of Linear Regression

Simple and Multiple Linear Regression

Automatic Dummy Variables creation technique

Model Validation parameters

Model Assumption testing

Splitting of data for Validation and testing

Business Case Study with real data to model in R software



Class 7: Linear Regression Practice Case Study

Participants will be asked to develop a Linear Regression model on a real life data, in presence of the instructor. Time given is 2.5 hours. Participants will be treated like an industry employee, but in terms of help certainly the instructor will not be as ruthless as the boss. After completion of the model (with the help of the instructor wherever it is required), the instructor will show how to present a model to a real life client.

Class 7: Logistic Regression

Predictive Analytics – Logistic Regression
Concepts of Logistic Regression
Difference between Linear Regression and Logistic Regression
Automatic Dummy Variables creation technique
Model Validation parameters
Model Assumption testing
Splitting of data for Validation and testing
Business Case Study with real data to model in R software

Class 8: Logistic Regression Practice Case Study

Participants will be asked to develop a Logistic Regression model on a real life data, in presence of the instructor. Time given is 2.5 hours. Participants will be treated like an industry employee, but in terms of help certainly the instructor will not be as ruthless as the boss. After completion of the model (with the help of the instructor wherever it is required), the instructor will show how to present a model to a real life client.

Class 9: Time Series Forecasting

Time series forecasting: ARIMA
Difference between forecasting and prediction
Concepts of time series data
Concepts of ARIMA
Descriptive analytics for ARIMA
Development of model
Best model selection
Forecasting with the best model
Residual analysis
Business Case Study with real data to model in R software
Participants will be asked to develop a model in presence of the instructor.



Class 10: Cluster Analysis

Unsupervised Machine Learning with R
Cluster Analysis: Concepts
Cluster analysis with R – K Means, Hierarchical etc.

Class 11: Decision Tree and Random Forest

Concepts of Decision Tree
Decision Tree with R
C5.0 algorithms and R part
Concepts of Random Forest
Random Forest with R

Important points:

1. After each class, assignments will be given as homework which are needed to be completed before the next class. The first 15 minutes of every class will be reserved to answer the participant's queries.
2. After every session, the discussed codes, presentations, handouts will be emailed to all the participants. Participants are advised to carry it either in soft copy or as print outs in the class.
3. Participants are advised to bring their own computers so that they can practice the codes along with the instructor.
4. Normally the class duration would be 3 hours, with a break of maximum 5-10 minutes depending of the requirement of the participants. In case all the queries of the participants are not answered within the stipulated time of 3 hours then the instructor will extend the class by 15 minutes to 30 minutes.
5. After the completion of the module, there will be an option for all the participants to work on other case studies on real life data for further practice. (This is optional and will not be considered for calculating your final grade)
6. If a participant feels that he/she requires further help on certain topic, then they can attend the same session of some other batch.