Mastering Data Analysis & Machine Learning with R: From Basic to Advanced

Course Outline

Certificate of Completion:
Only Basic part: ***"Basic Data Analysis & Machine Learning with R Programming"***
Basic & Advanced part: ***"Advanced Data Analysis & Machine Learning with R Programming"***

Contents		
Basic part		
Day 1	**Chapter 1**	**Introduction to Computer Programming and R** • Computer programming (basic concept) • Limited capabilities of computer. • Sample computer program using basic steps. • Variable, function & package concept. • Programming structures. • Introducing R Programming • R Advantages and Disadvantages • Environment setup • Installation of R • Installation of RStudio • Run some basic code.
Day 2	**Chapter 2**	**Core Concepts of Data Types, Structures, and Control Flow** • Data types • Data Structure • Variable, Keywords • Operators, Comment • Input from user • Data Structure in details • Vectors, Lists, Array • Matrix, Data frames, Factors • R Statements (If, if-else, else if, switch) • R Loops (For loop, repeat loop, while loop) • R Functions • Built-in functions
Day 3	**Chapter 3**	**R Packages and Repositories: Installation and Usage** • Packages in R Programming • What are repositories? • Install an R-Packages • Update, remove and check installed packages • Installing packages using RStudio UI • Load packages in R

		• Difference between a package and a library • Load more than one package at a time • Choose the right R Packages • Using package- gapminder
Day 4	**Chapter 4**	**Data Transformation and Manipulation** • Uses dplyr. • File Handling. • Creating a file, write into a file, rename a file, check existence of a file, reading a file, list all files, copy a file, create a directory. • Reading a text file, using readr package, read tabular data. • Working with CSV, Excel, JSON, XML file.
Day 5	**Chapter 5**	**Visualization Using R** • Basic ploting, Heatmap. • ggplot2 (grammar of graphics), Tidyverse. • Pie chart, Bar chart, Boxplot. • Histogram, Line graph, Scatterplot.
Day 6	**Chapter 6**	**Descriptive Statistics** • Central tendency: mean, median, mode. • Variability analysis or Dispersion: range, variance, standard deviation, minimum, maximum. • Normality test • Outlier detection • Create t-score and z-score • Percentile Score
Day 7	**Chapter 7**	**Correlation & Regression** • Definition of correlation • Pearson correlation, Spearman correlation, Kendall's tau • Partial correlation • Definition of regression • Types of regression • Simple linear regression • Linear bi-variate regression • Multiple linear regression • Non-linear regression • Logistic regression • Multinomial logistic regression
Day 8	**Chapter 8**	**Statistical Inference and Tests** • Some definitions of related terms. • Distinguish parametric and non-parametric test. • Crosstabs analysis • t-test and z-test • ANOVA • Median Test

		• Chi-square • Mann-Whitney U Test, K-S test • Wilcoxon Rank Test • Kruskal-Wallis
Day 9	**Chapter 9**	**Introduction to ML: Concepts, Types, & Model Evaluation** • What is machine learning? • Why machine learning? • When should you use machine learning? • Types of Systems of Machine Learning • Supervised and unsupervised learning • Supervised Learning • The most important supervised algorithms • Unsupervised Learning • Data Preprocessing ➢ Bad and Insufficient Quantity of Training Data ➢ Poor-Quality Data • Measures of Performance • Confusion Matrix

Advanced Part		
Day 10	**Chapter 10**	**Advanced Statistical Tests & Correlation** • Cronbach's alpha • Post Hoc tests • Tests for Measuring strength- phi, Cramér's V, etc. • Testing heteroscedasticity • Heteroscedasticity adjusted standard errors • Zero order vs Partial vs Semi-partial correlation • CI of correlation • Average correlation
Day 11	**Chapter 11**	**Advanced Regression & ML Concept** • Robust regression • Ridge Regression • Lasso Regression • Advanced unsupervised algorithms • Model-based learning
Day12	**Chapter 12**	**Survival & Time Series Analysis with Forecasting Techniques** • Life table • Kaplan-Meier • Cox regression • Related terms and definitions of Time Series Analysis • Decision build, analysis, forecasting

Day 13	**Chapter 13**	**Advanced Dimension Reduction, Classify and Clustering** • Principal component analysis (PCA) • Factor Analysis • Two step cluster • K-means cluster • Hierarchical cluster • Decision Tree • Linear Discriminant Analysis (LDA)
Day 14	**Chapter 14**	**Advanced Data Preprocessing & Model Evaluation Techniques** • Advanced Data Preprocessing ➢ Irrelevant Features ➢ Feature Engineering ➢ Testing ➢ Overfitting the Data ➢ Underfitting the Data • Recall • Recall Tradeoff • ROC
Day 15	**Chapter 15**	**Advanced Machine Learning Algorithms & Ensemble Methods** • Support Vector Machine (SVM) • Ensemble techniques • Bagging • Random Forests • Boosting ➢ Gradient Descent ➢ AdaBoost ➢ XGBoost