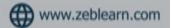# Data Science

"Change is the end result of all true learning."

ZebLearn is an ISO 9001-2015 Certified Company that is co-founded by highly experienced industry professionals and alumni of top universities. It is headquartered at Noida & It is one of the fastest-growing solution providers in the field of Education, IT, Consulting and Corporate Trainings.

Data Science Syllabus

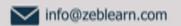## Introduction to Data Science with RPreview

- ❖ What is Data Science?
- ❖ Significance of Data Science in today's data-driven world, applications of Data Science, lifecycle of Data Science, and its components
- ❖ Introduction to Big Data Hadoop, Machine Learning, and Deep Learning
- ❖ Introduction to R programming and RStudio

Data Exploration

- ❖ Introduction to data exploration
- ❖ Importing and exporting data to/from external sources
- ❖ What are data exploratory analysis and data importing?
- ❖ DataFrames, working with them, accessing individual elements, vectors, factors, operators, in-built functions, conditional and looping statements, user-defined functions, and data types

Data ManipulationPreview

- ❖ Need for data manipulation
- ❖ Introduction to the dplyr package
- ❖ Selecting one or more columns with select(), filtering records on the basis of a condition with filter(), adding new columns with mutate(), sampling, and counting
- ❖ Combining different functions with the pipe operator and implementing SQL-like operations with sqldf
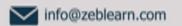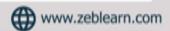
Data Science Syllabus

Data Visualization

- ❖ Introduction to visualization
- ❖ Different types of graphs, the grammar of graphics, the ggplot2 package, categorical distribution with geom_bar(), numerical distribution with geom_hist(), building frequency polygons with geom_freqpoly(), and making a scatterplot with geom_pont()
- ❖ Multivariate analysis with geom_boxplot
- ❖ Univariate analysis with a barplot, a histogram and a density plot, and multivariate distribution
- ❖ Creating barplots for categorical variables using geom_bar(), and adding themes with the theme() layer
- ❖ Visualization with plotly, frequency plots with geom_freqpoly(), multivariate distribution with scatter plots and smooth lines, continuous distribution vs categorical distribution with box-plots, and sub grouping plots
- ❖ Working with co-ordinates and themes to make graphs more presentable, understanding plotly and various plots, and visualization with ggvis
- ❖ Geographic visualization with ggmap() and building web applications with shinyR

Introduction to Statistics Preview

- ❖ Why do we need statistics?
- ❖ Categories of statistics, statistical terminology, types of data, measures of central tendency, and measures of spread
- ❖ Correlation and covariance, standardization and normalization, probability and the types, hypothesis testing, chi-square testing, ANOVA, normal distribution, and binary distribution
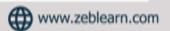
Data Science Syllabus

Machine Learning

- ❖ Introduction to Machine Learning
- ❖ Introduction to linear regression, predictive modeling, simple linear regression vs multiple linear regression, concepts, formulas, assumptions, and residuals in Linear Regression, and building a simple linear model
- ❖ Predicting results and finding the p-value and an introduction to logistic regression
- ❖ Comparing linear regression with logistics regression and bivariate logistic regression with multivariate logistic regression
- ❖ Confusion matrix the accuracy of a model, understanding the fit of the model, threshold evaluation with ROCR, and using qqnorm() and qqline()
- ❖ Understanding the summary results with null hypothesis, F-statistic, and
- ❖ building linear models with multiple independent variables

Logistic Regression Preview

- ❖ Introduction to logistic regression
- ❖ Logistic regression concepts, linear vs logistic regression, and math behind logistic regression
- ❖ Detailed formulas, logit function and odds, bivariate logistic regression, and Poisson regression
- ❖ Building a simple binomial model and predicting the result, making a confusion matrix for evaluating the accuracy, true positive rate, false positive rate, and threshold evaluation with ROCR
- ❖ Finding out the right threshold by building the ROC plot, cross validation, multivariate logistic regression, and building logistic models with multiple independent variables
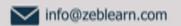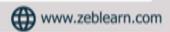- ❖ Real-life applications of logistic regression

Data Science Syllabus

Decision Trees and Random Forest

- ❖ What is classification? Different classification techniques
- ❖ Introduction to decision trees
- ❖ Algorithm for decision tree induction and building a decision tree in R
- ❖ Confusion matrix and regression trees vs classification trees
- ❖ Introduction to bagging
- ❖ Random forest and implementing it in R
- ❖ What is Naive Bayes? Computing probabilities
- ❖ Understanding the concepts of Impurity function, Entropy, Gini index, and Information gain for the right split of node
- ❖ Overfitting, pruning, pre-pruning, post-pruning, and cost-complexity pruning, pruning a decision tree and predicting values, finding out the right number of trees, and evaluating performance metrics

Unsupervised Learning Preview

- ❖ What is Clustering? Its use cases
- ❖ what is k-means clustering? What is canopy clustering?
- ❖ What is hierarchical clustering?
- ❖ Introduction to unsupervised learning
- ❖ Feature extraction, clustering algorithms, and the k-means clustering algorithm
- ❖ Theoretical aspects of k-means, k-means process flow, k-means in R, implementing k-means, and finding out the right number of clusters using a scree plot
- ❖ Dendograms, understanding hierarchical clustering, and implementing it in R
- ❖ Explanation of Principal Component Analysis (PCA) in detail and implementing PCA in R

Data Science Syllabus

Association Rule Mining and Recommendation Engines

❖ Introduction to association rule mining and MBA

❖ Measures of association rule mining: Support, confidence, lift, and apriori algorithm, and implementing them in R

❖ Introduction to recommendation engines

❖ User-based collaborative filtering and item-based collaborative filtering, and implementing a recommendation engine in R

❖ Recommendation engine use cases

Introduction to Artificial Intelligence

❖ Introducing Artificial Intelligence and Deep Learning

❖ What is an artificial neural network? TensorFlow: The computational framework for building AI models

❖ Fundamentals of building ANN using TensorFlow and working with TensorFlow in R
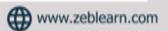
Time Series Analysis

❖ What is a time series? The techniques, applications, and components of time series

❖ Moving average, smoothing techniques, and exponential smoothing

❖ Univariate time series models and multivariate time series analysis

❖ ARIMA model

❖ Time series in R, sentiment analysis in R (Twitter sentiment analysis), and text analysis

Support Vector Machine (SVM)

❖ Introduction to Support Vector Machine (SVM)

❖ Data classification using SVM

❖ SVM algorithms using separable and inseparable cases

❖ Linear SVM for identifying margin hyperplane

Naïve Bayes

- ❖ What is the Bayes theorem?
- ❖ What is Naïve Bayes Classifier?
- ❖ Classification Workflow
- ❖ How Naive Bayes classifier works and classifier building in Scikit-Learn
- ❖ Building a probabilistic classification model using Naïve Bayes and the zero probability problem

Text MiningPreview

- ❖ Introduction to the concepts of text mining
- ❖ Text mining use cases and understanding and manipulating the text with 'tm' and 'stringR'
- ❖ Text mining algorithms and the quantification of the text

Thanks you

Now or Never