simplilearn

Data Science with R

Become an expert in data analytics using the R programming language in this data science certification training course. You'll master data exploration, data visualization, predictive analytics and descriptive analytics techniques with the R language. With this data science course, you'll get hands-on practice by implementing various real-life, industry-based projects in the domains of healthcare, retail, insurance, and many more.

About the Program

The Data Science Certification with R has been designed to give you in-depth knowledge of the various data analytics techniques that can be performed using R. The data science course is packed with real-life projects and case studies.

- Mastering R language: The data science course provides an in-depth understanding of the R language, R-studio, and R packages. You will learn the various types of apply functions including DPYR, gain an understanding of data structure in R, and perform data visualizations using the various graphics available in R.
- Mastering advanced statistical concepts: The data science training course also includes various statistical concepts such as linear and logistic regression, cluster analysis and forecasting.

Who Can Enroll For This Program?

There is an increasing demand for skilled data scientists across all industries, making this data science certification course well-suited for participants at all levels of experience. We recommend this Data Science training particularly for the following professionals:

IT professionals

Analytics professionals

Software Developers

Program features:

64 hours of in-depth learning

10 real-life industry-based projects

Lifetime access to self-paced learning

Dedicated mentoring session from our faculty of industry experts

Data Science with R Outcomes:

This data science training course will enable you to:

- Gain a foundational understanding of business analytics
- Install R, R-studio, and workspace setup, and learn about the various R packages
- Master R programming and understand how various statements are executed in R
- Gain an in-depth understanding of data structure used in R and learn to import/export data in R
- Define, understand and use the various apply functions and DPYR functions
- Understand and use the various graphics in R for data visualization
- Gain a basic understanding of various statistical concepts
- Understand and use hypothesis testing method to drive business decisions
- Understand and use linear, non-linear regression models, and classification techniques for data analysis
- Learn and use the various association rules and Apriori algorithm
- Learn and use clustering methods including K-means, DBSCAN, and hierarchical clustering

Chapter level details:

Lesson 1: Introduction to Business Analytics

Lesson Objective: This introductory lesson gives an overview of Business Analytics, it's types and applications.

Topics:

1. Overview
2. Business Decisions and Analytics
3. Types of Business Analytics
4. Applications of Business Analytics
5. Data Science Overview
6. Conclusion
7. Knowledge Check

Lesson 2: Introduction to R Programming

Lesson Objective: This lesson describes the importance of R, data types, variables, operators, conditional statements, loops, functions in R language.

Topics:

1. Overview
2. Importance of R
3. Data Types and Variables in R
4. Operators in R
5. Conditional Statements in R
6. Loops in R
7. R script
8. Functions in R
9. Conclusion
10. Knowledge Check

Lesson 3: Data Structures

Lesson Objective: This lesson teaches you to identify data structures, assign values to data structures, manipulate data by assigning values and applying functions.

Topics:

1. Overview
2. Identifying Data Structures
3. Demo: Identifying Data Structures
4. Assigning Values to Data Structures
5. Data Manipulation
6. Demo: Assigning values and applying functions
7. Conclusion
8. Knowledge Check

Lesson 4: Data Visualization

Lesson Objective: This lesson will teach you how to visualize data using graphics in R and to plot the outputs.

Topics:

1. Overview
2. Introduction to Data Visualization
3. Data Visualization using Graphics in R
4. ggplot2
5. File Formats of Graphic Outputs
6. Conclusion
7. Knowledge Check

Lesson 5: Statistics for Data Science-I

Lesson Objective: This lesson will introduce you to hypothesis and data sampling techniques.

Topics:

1. Overview
2. Introduction to Hypothesis
3. Types of Hypothesis
4. Data Sampling
5. Confidence and Significance Levels
6. Conclusion
7. Knowledge Check

Lesson 6: Statistics for Data Science-II

Lesson Objective: This lesson will give you a detailed overview of the Hypothesis, Parametric, and Non-parametric tests.

Topics:

1. Overview
2. Hypothesis Test
3. Parametric Test
4. Non-Parametric Test
5. Hypothesis Tests about Population Means
6. Hypothesis Tests about Population Variance
7. Hypothesis Tests about Population Proportions
8. Conclusion
9. Knowledge Check

Lesson 7: Regression Analysis

Lesson Objective: Regression Analysis overview, types of regression Analysis, and regression models are covered in this lesson.

Topics:

1. Overview
2. Introduction to Regression Analysis
3. Types of Regression Analysis Models
4. Linear Regression
5. Demo: Simple Linear Regression
6. Non-Linear Regression
7. Demo: Regression Analysis with Multiple Variables

8. Cross Validation
9. Non-Linear to Linear Models
10. Principal Component Analysis
11. Factor Analysis
12. Conclusion
13. Knowledge Check

Lesson 8: Classification

Lesson Objective: Logistic regression, support vector machines, naive bayes classifier, decision tree classification, random forest classification, classifier models, K-fold cross validation are covered in this lesson.

Topics:

1. Overview
2. Classification and Its Types
3. Logistic Regression
4. Support Vector Machines
5. Demo: Support Vector Machines
6. K-Nearest Neighbours
7. Naive Bayes Classifier
8. Demo: Naive Bayes Classifier
9. Decision Tree Classification
10. Demo Decision Tree Classification
11. Random Forest Classification
12. Evaluating Classifier Models
13. Demo: K-Fold Cross Validation
14. Conclusion
15. Knowledge Check

Lesson 9: Clustering

Lesson Objective: This lesson covers clustering, clustering methods, and types of clustering: K-means and Hierarchical clustering.

Topics:

1. Overview
2. Introduction to Clustering
3. Clustering Methods
4. Demo: K-means Clustering
5. Demo: Hierarchical Clustering
6. Conclusion
7. Knowledge Check

Lesson 10: Association

Lesson Objective: This concluding lesson covers association rules and apriori algorithm.

Topics:

1. Overview
2. Association Rule
3. Apriori Algorithm
4. Demo: Apriori Algorithm
5. Conclusion
6. Knowledge Check

Projects Covered:

The data science certification course includes ten real-life, industry-based projects. Successful evaluation of one of the following six projects is a part of the certification eligibility criteria.

Project 1: Products rating prediction for Amazon

Amazon, one of the leading US-based e-commerce companies, recommends products within the same category to customers based on their activity and reviews on other similar products. Amazon would like to improve this recommendation engine by predicting ratings for the non-rated products and add them to recommendations accordingly.

Domain: E-commerce

Project 2: Demand Forecasting for Walmart

Predict accurate sales for 45 stores of Walmart, one of the US-based leading retail stores, considering the impact of promotional markdown events. Check if macroeconomic factors like CPI, unemployment rate, etc. have an impact on sales.

Domain: Retail

Project 3: Improving customer experience for Comcast

Comcast, one of the US-based global telecommunication companies wants to improve customer experience by identifying and acting on problem areas that lower customer satisfaction if any. The company is also looking for key recommendations that can be implemented to deliver the best customer experience.

Domain: Telecom

Project 4: Attrition Analysis for IBM

IBM, one of the leading US-based IT companies, would like to identify the factors that influence attrition of employees. Based on the parameters identified, the company would also like to build a logistics regression model that can help predict if an employee will churn or not.

Domain: Workforce Analytics

Project 5:

A nationwide survey of hospital costs conducted by the US Agency for Healthcare consists of hospital records of inpatient samples. The given data is restricted to the city of Wisconsin and relates to patients in the age group 0-17 years. The agency wants to analyze the data to research on the health care costs and their utilization.

Domain: Healthcare

Project 6:

The data gives the details of third party motor insurance claims in Sweden for the year 1977. In Sweden, all motor insurance companies apply identical risk arguments to classify customers, and thus their portfolios and their claims statistics can be combined. The data were compiled by a Swedish Committee on the Analysis of Risk Premium in Motor Insurance. The Committee was asked to look into the problem of analyzing the real influence on the claims of the risk arguments and to compare this structure with the actual tariff.

Domain: Insurance

Project 7:

A high-end fashion retail store is looking to expand its products. It wants to understand the market and find the current trends in the industry. It has a database of all products with attributes, such as style, material, season, and the sales of the products over a period of two months.

Domain: Retail

Project 8:

The web analytics team of www.datadb.com is interested to understand the web activities of the site, which are the sources used to access the website. They have a database that states the keywords of time in the page, source group, bounces, exits, unique page views, and visits.

Domain: Internet

Project 9:

An education department in the US needs to analyze the factors that influence the admission of a student into a college. Analyze the historical data and determine the key drivers.

Domain: Education

Project 10:

A UK-based online retail store has captured the sales data for different products for the period of one year (Nov 2016 to Dec 2017). The organization sells gifts primarily on the online platform. The customers who make a purchase consume directly for themselves. There are small businesses that buy in bulk and sell to other customers through the retail outlet channel. Find significant customers for the business who make high purchases of their favourite products.

Domain: E-commerce

The course also includes 4 more projects for you to practice.

Project 11:

Details of listener preferences are recorded online. This data is not only used for recommending music that the listener is likely to enjoy but also to drive a focused marketing strategy that sends out advertisements for music that a listener may wish to buy. Using the demographic data, predict the music preferences of the user for targeted advertising.

Domain: Music Industry

Project 12:

You'll predict whether someone will default or not default on a loan based on user demographic data. You'll perform logistic regression by considering the loan's features and the characteristics of the borrower as explanatory variables.

Domain: Finance

Project 13:

Analyze the monthly, seasonally-adjusted unemployment rates for U.S. employment data of all 50 states, covering the period from January 1976 through August 2010. The requirement is to cluster the states into groups that are alike using a feature vector.

Domain: Unemployment

Project 14:

Flight delays are frequently experienced when flying from the Washington DC area to the New York City area. By using logistical regression, you'll identify flights that are likely to be delayed. The provided dataset helps with a number of variables including airports and flight times.

Domain: Airline