



Data Science Certification Training

Table of Contents

1. About the Program
2. About Intellipaap
3. Key Features
4. Career Support
5. Why take up this course?
6. Who should take up this course?
7. Program Curriculum
8. Self-paced Courses
9. Project Work
10. Certification
11. Intellipaap Success Stories
12. Contact Us





Data Scientist is the best job in the
21st century – Harvard Business Review

About the Program

This Data Scientist course in collaboration with IBM offers you multiple courses on Data Science to let you master skills such as data analytics, R programming, statistical computing, Machine Learning algorithms, k-means clustering, and more. It includes multiple hands-on exercises and project work in the domains of banking, finance, entertainment, etc. Intellipaat’s online Data Science courses are well recognized across 500+ employers helping you to land in your dream job.

About Intellipaat

Intellipaat is one of the leading e-learning training providers with more than 600,000 learners across 55+ countries. We are on a mission to democratize education as we believe that everyone has the right to quality education. Our courses are delivered by subject matter experts from top MNCs, and our world-class pedagogy enables learners to quickly learn difficult topics in no time. Our 24/7 technical support and career services will help them jump-start their careers in their dream companies.

Key Features



**42 HRS INSTRUCTOR-LED
TRAINING**



28 HRS SELF-PACED TRAINING



**56 HRS REAL-TIME
PROJECT WORK**



LIFETIME ACCESS



24/7 TECHNICAL SUPPORT



**INDUSTRY-RECOGNIZED
CERTIFICATION**



**JOB ASSISTANCE THROUGH
80+ CORPORATE TIE-UPS**



FLEXIBLE SCHEDULING

Career Support



SESSIONS WITH INDUSTRY MENTORS

Attend sessions from top industry experts and get guidance on how to boost your career growth



MOCK INTERVIEWS

Mock interviews to make you prepare for cracking interviews by top employers



GUARANTEED INTERVIEWS & JOB SUPPORT

Get interviewed by our 400+ hiring partners



RESUME PREPARATION

Get assistance in creating a world-class resume from our career services team



Why take up this course?

The average annual salary of Data Scientists as per [Indeed](#) is US\$122,801 in the United States.

- Data Scientist is the best job in the 21st century – Harvard Business Review
- The number of jobs for all data professionals in the United States will increase to 2.7 million – IBM
- Global Big Data market achieves US\$122 billion in sales in 6 years – Frost & Sullivan

The demand for Data Scientists far exceeds the supply. This is a serious problem in the data-driven world that we are living in today. As a result, most organizations are willing to pay high salaries for professionals with appropriate Data Science skills.

Data science training online will help you become proficient in Data Science, R programming language, Data Analysis, Big Data, and more. Thus, you can easily accelerate your career in this evolving domain and take it to the next level.

Who should take up this course?

- Information Architects and Statisticians
- Developers looking to master Machine Learning and Predictive Analytics
- Big Data experts, Business Analysts, and Business Intelligence professionals
- Aspirants looking to work as Machine Learning experts, Data Scientists, etc.

Program Curriculum

DATA SCIENCE COURSE CONTENT

1. INTRODUCTION TO DATA SCIENCE WITH R

1.1 What is Data Science?

1.2 Significance of Data Science in today's data-driven world, its applications, lifecycle, and its components

1.3 Introduction to R programming and RStudio

Hands-on Exercise: *Installation of RStudio and implementing simple mathematical operations and logic using R operators, loops, if statements, and switch cases*

2. DATA EXPLORATION

2.1 Introduction to data exploration

2.2 Importing and exporting data from/to external sources

2.3 What are data exploratory analysis and data importing?

2.4 DataFrames, working with them, accessing individual elements, vectors, factors, operators, built-in functions, conditional and looping statements, user-defined functions, and data types

Hands-on Exercise: *Accessing the individual elements of customer churn data and modifying and extracting the results from a dataset using user-defined functions in R*

3. DATA MANIPULATION

3.1 Need for data manipulation

3.2 Introduction to the dplyr package

3.3 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

3.4 Combining different functions with the pipe operator and implementing SQL-like operations with `sqldf`

Hands-on Exercise: *Implementing dplyr, performing various operations for manipulating data, and storing it*

4. DATA VISUALIZATION

4.1 Introduction to visualization

4.2 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_point()`

4.3 Multivariate analysis with `geom_boxplot`

4.4 Univariate analysis with a bar plot, a histogram, and a density plot and multivariate distribution

4.5 Creating bar plots for categorical variables using `geom_bar()` and adding themes with the `theme()` layer

4.6 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

4.7 Working with co-ordinates and themes to make graphs more presentable, understanding plotly and various plots, and visualization with ggvis

4.8 Geographic visualization with `ggmap()` and building web applications with shinyR

Hands-on Exercise: *Creating data visualizations to understand the customer churn ratio using charts with ggplot2, using plotly for importing and analyzing data into grids, and visualizing tenure, monthly charges, total charges, and other individual columns by using a scatter plot*

5. INTRODUCTION TO STATISTICS

5.1 Why do we need statistics?

5.2 Categories of statistics, statistical terminology, types of data, measures of central tendency, and measures of spread

5.3 Correlation and covariance, standardization and normalization, probability and the types, hypothesis testing, chi-square testing, ANOVA, normal distribution, and binary distribution

Hands-on Exercise: *Building a statistical analysis model that uses quantifications, representations, and experimental data for gathering, reviewing, analyzing, and drawing conclusions from the data*

6. MACHINE LEARNING

6.1 Introduction to Machine Learning

6.2 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

6.3 Predicting results, finding the p-value, and an introduction to logistic regression

6.4 Comparing linear regression with logistics regression and bivariate logistic regression with multivariate logistic regression

6.5 Confusion matrix for checking the accuracy of a model, understanding the fit of the model, threshold evaluation with ROCR, and using `qqnorm()` and `qqline()`

6.6 Understanding the summary results with null hypothesis, F-statistic, and building linear models with multiple independent variables

Hands-on Exercise: Modeling relationships within data using linear predictor functions and implementing linear and logistic regression in R by building a model with 'tenure' as the dependent variable and with multiple independent variables

7. LOGISTIC REGRESSION

7.1 Introduction to logistic regression

7.2 Logistic regression concepts, linear vs logistic regression, and math behind logistic regression

7.3 Detailed formulas, logit function and odds, bivariate logistic regression, and Poisson regression

7.4 Building a simple binomial model, predicting the result, making a confusion matrix for evaluating the accuracy, true positive rate, false positive rate, and threshold evaluation with ROCR

7.5 Finding out the right threshold by building the ROC plot, cross validation, multivariate logistic regression, and building logistic models with multiple independent variables

7.6 Real-life applications of logistic regression

Hands-on Exercise: Implementing predictive analytics by describing data, explaining the relationship between one dependent binary variable and one or more binary variables, and using `glm()` to build a model with 'Churn' as the dependent variable

8. DECISION TREES AND RANDOM FOREST

8.1 What is classification? Different classification techniques

8.2 Introduction to decision trees

8.3 Algorithm for decision tree induction and building a decision tree in R

8.4 Confusion matrix and regression trees vs classification trees

8.5 Introduction to bagging

8.6 Random forest and implementing it in R

8.7 What is Naive Bayes? Computing probabilities

8.8 Understanding the concepts of Impurity function, Entropy, Gini index, and Information gain for the right split of node

8.9 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

Hands-on Exercise: *Implementing a random forest for both regression and classification problems, building a tree, pruning it by using 'churn' as the dependent variable, building a random forest with the right number of trees, and using ROC for performance metrics*

9. UNSUPERVISED LEARNING

9.1 What is Clustering? Its use cases

9.2 what is k-means clustering? What is canopy clustering?

9.3 What is hierarchical clustering?

9.4 Introduction to unsupervised learning

9.5 Feature extraction, clustering algorithms, and the k-means clustering algorithm

9.6 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

9.7 Dendograms, understanding hierarchical clustering, and implementing it in R

9.8 Explanation of Principal Component Analysis (PCA) in detail and implementing it in R

Hands-on Exercise: *Deploying unsupervised learning with R to achieve clustering and dimensionality reduction and using k-means clustering for visualizing and interpreting results for the customer churn data*

10. ASSOCIATION RULE MINING AND RECOMMENDATION ENGINE

10.1 Introduction to association rule mining and MBA

10.2 Measures of association rule mining: Support, confidence, lift, and the Apriori algorithm, and implementing them in R

10.3 Introduction to recommendation engines

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

10.5 Recommendation engine use cases

Hands-on Exercise: Deploying association analysis as a rule-based Machine Learning method and identifying strong rules discovered in databases with measures based on interesting discoveries

Self-paced Course Content

1. INTRODUCTION TO ARTIFICIAL INTELLIGENCE

- 11.1 Introducing Artificial Intelligence and Deep Learning
- 11.2 What is an artificial neural network? TensorFlow: The computational framework for building AI models
- 11.3 Fundamentals of building ANN using TensorFlow and working with TensorFlow in R

2. TIME SERIES ANALYSIS

- 12.1 What is a time series? The techniques, applications, and components of time series
- 12.2 Moving average, smoothing techniques, and exponential smoothing
- 12.3 Univariate time series models and multivariate time series analysis
- 12.4 ARIMA model
- 12.5 Time series in R, sentiment analysis in R (Twitter sentiment analysis), and text analysis

Hands-on Exercise: Analyzing time series data and checking the sequence of measurements that follow a non-random order to identify the nature of phenomenon and to forecast the future values in the series

3. SUPPORT VECTOR MACHINE (SVM)

- 13.1 Introduction to Support Vector Machine (SVM)
- 13.2 Data classification using SVM
- 13.3 SVM algorithms using separable and inseparable cases
- 13.4 Linear SVM for identifying margin hyperplane

4. NAÏVE BAYES

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

5. TEXT MINING

15.1 Introduction to the concepts of text mining

15.2 Text mining use cases and understanding and manipulating the text with 'tm' and 'stringR'

15.3 Text mining algorithms and the quantification of the text

15.4 TF-IDF and after TF-IDF

Project Work

Market Basket Analysis

This is an inventory management project where you will find the trends in the data that will help the company to increase sales. In this project, you will be implementing association rule mining, data extraction, and data manipulation for the market basket analysis.

Credit Card Fraud Detection

The project consists of data analysis for various parameters of a banking dataset. You will be using a V7 predictor, a V4 predictor for analysis, and data visualization for finding the probability of occurrence of fraudulent activities.

Loan Approval Prediction

In this project, you will use the banking dataset for data analysis, data cleaning, data preprocessing, and data visualization. You will implement algorithms such as Principal Component Analysis and Naive Bayes after data analysis to predict the approval rate of a loan using various parameters.

Netflix Recommendation System

Here, you will implement exploratory data analysis, data manipulation, and visualization to understand and find the trends in the Netflix dataset. You will use various Machine Learning algorithms such as association rule mining, classification algorithms, and many more to create movie recommendation systems for viewers using the Netflix dataset.

Case Study 1: Introduction to R Programming

In this project, you need to work with several operators involved in R programming, including relational operators, arithmetic operators, and logical operators, for various organizational needs.

Case Study 2: Solving Customer Churn Using Data Exploration

Using data exploration, you have to understand what needs to be done to make reductions in customer churn. In this project, you will be required to extract individual columns, use loops to work on repetitive operations, and create and implement filters for data manipulation.

Case Study 3: Creating Data Structures in R

You will implement numerous data structures for possible scenarios. This project requires you to create and use vectors. Further, you need to build and use metrics, utilize arrays for storing those metrics, and have knowledge of lists.

Case Study 4: Implementing SVD in R

In this project, you will utilize the dataset of MovieLens to analyze and understand single value decomposition and its use in R programming. Further, you must build custom recommended movie sets for all users and develop a collaborative filtering model based on the users. Also, for movie recommendation, you must create `realRatingMatrix`.

Case Study 5: Time Series Analysis

This project required you to perform TSA and understand ARIMA and its concepts with respect to a given scenario. Here, you will use the R programming language, the ARIMA model, time series analysis, and data visualization. You must know how to build an ARIMA model and fit it, find optimal parameters by plotting PACF charts, and perform various analyses to predict values.

Certification

After the completion of the course, you will get a certificate from IntelliPaat.



Success Stories



Kevin K Wada

Thank you very much for your top-class service. A special mention should be made for your patience in listening to my queries and giving me a solution, which was exactly what I was looking for. I am giving you a 10 on 10!



Shreyashkumar Limbhetwala

I want to talk about the rich LMS that Intellipaat's Data Science program offered. The extensive set of PPTs, PDFs, and other related material were of the highest quality, and due to this, my learning with Intellipaat was excellent. I could clear the Data Scientist certification exam in the first attempt.



Sudipto

Intellipaat's Data Scientist training is outstanding. The trainer is an experienced Data Scientist who has a good hold of the subject. Now, I am an expert in Data Science, and I am already placed in a reputed firm.



Afsana Zaman

Great experience! The support team has always been supportive and efficient to help me. The collaboration of practical exposure with theoretical knowledge makes Intellipaat highly suitable for those who want to upgrade their career.



Shamirna Micheal

It is a genuine platform for learning. I finished my course recently from Intellipaat. The trainers were excellent at teaching. Further, the course was well-structured, and the lectures were really flexible. I am currently working, and I still got the time to complete the course within the given time. It was mainly possible because of the 24/7 support system and the clarity of their teaching.

CONTACT US

INTELLIPAAT SOFTWARE SOLUTIONS PVT. LTD.

Bangalore

AMR Tech Park 3, Ground Floor, Tower B,
Hongasandra Village, Bommanahalli,
Hosur Road, Bangalore – 560068

USA

1219 E. Hillsdale Blvd. Suite 205,
Foster City, CA 94404

If you have any further queries or just want to have a conversation with us, then do call us.

IND: +91-7022374614 | US: 1-800-216-8930