









Data Science Training Course



About IntelliPaat

IntelliPaat is a fast-growing professional training provider that is offering training in over 150 most sought-after tools and technologies. We have a learner base of 600,000 in over 32 countries and growing. For job assistance and placement we have direct tie-ups with 80+ MNCs.

Key Features of IntelliPaat Training:

 Instructor Led Training 42 Hrs of highly interactive instructor led training	 Self-Paced Training 28 Hrs of Self-Paced session with Lifetime access	 Exercise and project work 56 Hrs of real-time projects after every module	 Lifetime Access Lifetime access and free upgrade to latest version
 Support Lifetime 24*7 technical support and query resolution	 Get Certified Get global industry recognized certifications	 Job Assistance Job assistance through 80+ corporate tie-ups	 Flexi Scheduling Attend multiple batches for lifetime & stay updated.

About the Course

This Data Scientist course online provides detailed learning through self-paced videos and live instructor-led sessions that help you gain skills in the shortest possible time. Data Scientists are among the highest-paid and most in demand professionals. This in-depth Data Scientist course covers 'What is Data Science?', statistical methods, data acquisition and analysis, Machine Learning algorithms, predictive analytics, data modeling, etc. At the end of the course, you will work on building a recommendation engine for an ecommerce site and will work on a real-time capstone project.

 Instructor Led Duration – 42 Hrs Weekend Batch –3 Hrs/Session	 Self Paced Duration – 28 Hrs
---	--

Why take this Course?

The average annual salary of Data Scientists as per Indeed is approximately 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 a 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.

Course Content

Module /Topic
<p>Module 01 - Introduction to Data Science with R</p> <ul style="list-style-type: none">❖ 1.1 What is Data Science?❖ 1.2 Significance of Data Science in today's data-driven world, applications of Data Science, lifecycle of Data Science, and its components❖ 1.3 Introduction to Big Data Hadoop, Machine Learning, and Deep Learning❖ 1.4 Introduction to R programming and RStudio <p>Hands-on Exercise:</p> <ul style="list-style-type: none">❖ 1. Installation of RStudio❖ 2. Implementing simple mathematical operations and logic using R operators, loops, if statements, and switch cases
<p>Module 02 - Data Exploration</p> <ul style="list-style-type: none">❖ 2.1 Introduction to data exploration❖ 2.2 Importing and exporting data to/from external sources❖ 2.3 What are data exploratory analysis and data importing?❖ 2.4 DataFrames, working with them, accessing individual elements, vectors, factors, operators, in-built functions, conditional and looping statements, user-defined functions, and data types <p>Hands-on Exercise:</p> <ul style="list-style-type: none">❖ 1. Accessing individual elements of customer churn data❖ 2. Modifying and extracting results from the dataset using user-defined functions in R

Module 03 - 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:

- ❖ 1. Implementing dplyr
- ❖ 2. Performing various operations for manipulating data and storing it

Module 04 - 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 barplot, a histogram and a density plot, and multivariate distribution
- ❖ 4.5 Creating barplots 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:

- ❖ 1. Creating data visualization to understand the customer churn ratio using `ggplot2` charts
- ❖ 2. Using `plotly` for importing and analyzing data
- ❖ 3. Visualizing tenure, monthly charges, total charges, and other individual columns using a scatter plot

Module 05 - 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:

- ❖ 1. Building a statistical analysis model that uses quantification, representations, and experimental data
- ❖ 2. Reviewing, analyzing, and drawing conclusions from the data

Module 06 - 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 and 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 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:

- ❖ 1. Modeling the relationship within data using linear predictor functions
- ❖ 2. Implementing linear and logistics regression in R by building a model with 'tenure' as the dependent variable

Module 07 - 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 and 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:

- ❖ 1. Implementing predictive analytics by describing data
- ❖ 2. Explaining the relationship between one dependent binary variable and one or more binary variables
- ❖ 3. Using glm() to build a model, with 'Churn' as the dependent variable

Module 08 - 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:

- ❖ 1. Implementing random forest for both regression and classification problems
- ❖ 2. Building a tree, pruning it using 'churn' as the dependent variable, and building a random forest with the right number of trees
- ❖ 3. Using ROCR for performance metrics

Module 09 - 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 PCA in R

Hands-on Exercise:

- ❖ 1. Deploying unsupervised learning with R to achieve clustering and dimensionality reduction
- ❖ 2. K-means clustering for visualizing and interpreting results for the customer churn data

Module 10 - Association Rule Mining and Recommendation Engines

- ❖ 10.1 Introduction to association rule mining and MBA
- ❖ 10.2 Measures of association rule mining: Support, confidence, lift, and 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:

- ❖ 1. Deploying association analysis as a rule-based Machine Learning method
- ❖ 2. Identifying strong rules discovered in databases with measures based on interesting discoveries

Data Science Training Course

Self-paced Course Content

Module 11 - 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

Module 12 - 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:

- ❖ 1. Analyzing time series data
- ❖ 2. Analyzing the sequence of measurements that follow a non-random order to identify the nature of phenomenon and forecast the future values in the series

Module 13 - 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

Hands-on Exercise:

- ❖ 1. Analyzing time series data
- ❖ 2. Analyzing the sequence of measurements that follow a non-random order to identify the nature of phenomenon and forecast the future values in the series

Module 14 - 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

Self-paced Course Content

Module 15 - 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

Case Studies

- ❖ Case Study 01: Market Basket Analysis (MBA)
 - ❖ 1.1 This case study is associated with the modeling technique of Market Basket Analysis, where you will learn about loading data, plotting items, and running algorithms.
 - ❖ 1.2 It includes finding out the items that go hand in hand and can be clubbed together.
 - ❖ 1.3 This is used for various real-world scenarios like a supermarket shopping cart and so on.
- ❖ Case Study 02: Logistic Regression
 - ❖ 2.1 In this case study, you will get a detailed understanding of the advertisement spends of a company that will help drive more sales.
 - ❖ 2.2 You will deploy logistic regression to forecast future trends.
 - ❖ 2.3 You will detect patterns and uncover insight using the power of R programming.
 - ❖ 2.4 Due to this, the future advertisement spends can be decided and optimized for higher revenues.
- ❖ Case Study 03: Multiple Regression
 - ❖ 3.1 You will understand how to compare the miles per gallon (MPG) of a car based on various parameters.
 - ❖ 3.2 You will deploy multiple regression and note down the MPG for car make, model, speed, load conditions, etc.
 - ❖ 3.3 The case study includes model building, model diagnostic, and checking the ROC curve, among other things.
- ❖ Case Study 04: Receiver Operating Characteristic (ROC)
 - ❖ 4.1 In this case study, you will work with various datasets in R.
 - ❖ 4.2 You will deploy data exploration methodologies.
 - ❖ 4.3 You will also build scalable models.
 - ❖ 4.4 Besides, you will predict the outcome with highest precision, diagnose the model that you have created with real-world data, and check the ROC curve.

Project Work

Project 01: Market Basket Analysis

Description: 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.

Project 02: Credit Card Fraud Detection

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

Project 03: Loan Approval Prediction

Description: 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.

Project 04: Netflix Recommendation System

Description: 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 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

Use data exploration in order 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

Implement numerous data structures for numerous 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

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

Case Study 5: Time Series Analysis

Description: 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, ARIMA model, time series analysis, and data visualization. So, you must understand how to build an ARIMA model and fit it, find optimal parameters by plotting PACF charts, and perform various analyses to predict values.

Data Science Training Course

Intellipaap Job Assistance Program

Intellipaap is offering comprehensive job assistance to all the learners who have successfully completed the training. A learner will be considered to have successfully completed the training if he/she finishes all the exercises, case studies, projects and gets a minimum of 60% marks in the Intellipaap qualifying exam.

Intellipaap has exclusive tie-ups with over 80 MNCs for placement. All the resumes of eligible candidates will be forwarded to the Intellipaap job assistance partners. Once there is a relevant opening in any of the companies, you will get a call directly for the job interview from that particular company.

Frequently Asked Questions:

Q 1. What is the criterion for availing the Intellipaap job assistance program?

Ans. All Intellipaap learners who have successfully completed the training post, April 2017 are directly eligible for the Intellipaap job assistance program.

Q 2. Which are the companies that I can get placed in?

Ans. We have exclusive tie-ups with MNCs like **Ericsson, Cisco, Cognizant, Sony, Mu Sigma, Saint-Gobain, Standard Chartered, TCS, Genpact, Hexaware**, and more. So you have the opportunity to get placed in these top global companies.

Q 3. Does Intellipaap help learners to crack the job interviews?

Ans. Intellipaap has an exclusive section which includes the top interview questions asked in top MNCs for most of the technologies and tools for which we provide training. Other than that our support and technical team can also help you in this regard.

Q 4. Do I need to have prior industry experience for getting an interview call?

Ans. There is no need to have any prior industry experience for getting an interview call. In fact, the successful completion of the Intellipaap certification training is equivalent to six months of industry experience. This is definitely an added advantage when you are attending an interview.

Q 5. What is the job location that I will get?

Ans. Intellipaap will try to get you a job in your same location provided such a vacancy exists in that location.

Q 6. Which is the domain that I will get placed in?

Ans. Depending on the Intellipaap certification training you have successfully completed, you will be placed in the same domain.

Q 7. Is there any fee for the Intellipaap placement assistance?

Ans. Intellipaap does not charge any fees as part of the placement assistance program.

Q 8. If I don't get a job in the first attempt, can I get another chance?

Ans. Definitely, yes. Your resume will be in our database and we will circulate it to our MNC partners until you get a job. So there is no upper limit to the number of job interviews you can attend.

Q 9. Does Intellipaap guarantee a job through its job assistance program?

Ans. Intellipaap does not guarantee any job through the job assistance program. However, we will definitely offer you full assistance by circulating your resume among our affiliate partners.

Q 10. What is the salary that I will be getting once I get the job?

Ans. Your salary will be directly commensurate with your abilities and the prevailing industry standards.

Data Science Training Course



What makes us who we are?



"I want to talk about the rich LMS that Intellipaate data science training offered. The extensive set of PPTs, PDFs, and other related Data Science course material were of the highest quality and due to this my learning with Intellipaate was excellent and I could clear the Cloud era Data Scientist certification in the first attempt."

-Shreyash Limbhetwala



"I had taken the data science master program which is a combo of SAS, R and Apache Mahout. Since there are so many technologies involved in the Data Science course, getting your query resolved at the right time becomes the most important aspect. But with Intellipaate there was no such problem as all my queries were resolved in less than 24 hours."

- Giri Karnal



"The Intellipaate data science certification training videos really made me excited about studying data science. They were so elaborate and so professionally created that I could learn data science from the comfort of my home thanks to those learner-friendly videos. I am grateful to Intellipaate. "

- Nitesh Kumar