

CERTIFIED DATA SCIENTIST

Using R | Duration : 5 Days

GET CERTIFIED

TRANSFORM YOUR CAREER

COURSE DESCRIPTION

The data scientist role is an offshoot of the statistician role that includes the use of advanced analytics technologies, including machine learning and predictive modeling, to provide insights beyond statistical analysis. The demand for data science skills has grown significantly in recent years as companies look to glean useful information from the voluminous amounts of structured, unstructured and semi-structured data that a large enterprise produces and collects – collectively referred to as big data.

A data scientist uses large amounts of data to develop hypotheses, make inferences and hone in on customer, business and market trends. The data scientist must be able to communicate how to use analytics data to drive business decisions that may include changing course, improving a process or product, or creating new services or products.

COURSE OBJECTIVE

This is a complete course that provides you detailed understanding of data science, encompasses basic statistical concepts to advanced analytics and predictive modeling techniques, along with project life cycle, data acquisition, analysis, statistical methods and machine learning. The objective of the course is to learn statistical analysis techniques and tool to solve business problems that help you to emerge as 'Industry Ready' professional in the field of Data Science. You will be learning the Data Science skills with the most popular and leading analytical tool widely used across industries such as R.

Upon completion of Data Science course, you now have acquired valuable skills to

1. Master key facets of data investigation, including data wrangling, cleaning, sampling, management, exploratory analysis, regression and classification, prediction, and data communication.
2. Implement foundational concepts of data computation, such as data structure, algorithms, parallel computing, simulation, and analysis.
3. Leverage your knowledge of key subject areas, such as statistical quality control, exponential smoothing, seasonally adjusted trend analysis, or data visualization.

WHO SHOULD ATTEND?

- Graduates / Professionals from various quantitative backgrounds like Engineering, Finance, Mathematics, Statistics, Business Management who aspires to spearhead their career in Data Analytics.
- Basic knowledge of data analysis & business problems
- Analytics consultants
- IT/Software Professionals

PREREQUISITES

- Prior knowledge in basic programming / statistics is recommended for this course.



WHY DATABYTE ?



HIGHLY QUALIFIED & EXPERIENCED TRAINERS



INDUSTRY RELEVANT TRAINING PROGRAM



FULLY HANDS-ON PRACTICAL WORKING ON REAL-LIFE DATA SETS AND CASE STUDIES



INSTRUCTOR-LED CLASSROOM BASED TRAINING



CRAFTED BY EXPERTS TO KEEP YOU AHEAD IN INDUSTRY BEST PRACTICES

INTRODUCTION TO DATA SCIENCE

- What is Data Science?
- Analytics vs. Data warehousing, OLAP, MIS Reporting
- Relevance in industry and need of the hour
- Types of problems and business objectives in various industries
- (Regression, classification, segmentation, forecasting, optimization etc)
- How leading companies are harnessing the power of analytics?
- Critical success drivers
- Different phases of a typical Analytics projects
- Understanding Heuristic vs. statistical models/analysis
- Understanding classical techniques vs. machine learning techniques
- Latest Trends in data science
- Opportunities with data science

INTRODUCTION TO R-ENVIRONMENT

- The Workspace
- Input/ Output
- Useful Packages (Base & other packages)
- Graphic User Interfaces
- Customizing Startup
- Batch Processing and reusing Results

DATA INPUT & OUTPUT (IMPORTING & EXPORTING)

- Data Structure & Data Types (Vectors, Matrices, factors, Data frames, and Lists)
- Importing Data from various sources
- Database Input (Connecting to database)
- Exporting Data to various formats
- Viewing Data (Viewing partial data and full data)
- Variable & Value Labels – Date Values

DATA MANIPULATION

- Creating New Variables (calculations & Binning)
- Operators (Using multiple operators)
- Built-in Functions & User Defined Function
- Control Structures (conditional statements, Loops, apply functions)
- Sorting Data
- Merging and Appending Data
- Aggregating/summarizing Data
- Reshaping Data
- Sub setting Data
- Data Type Conversions
- Sampling
- Renaming-formatting data
- Handling duplicates/Missing

PROJECT - CASE STUDIES

- Key Drivers for Customer Spending.
- Prediction bad Customers (default Customers) using Credit Customer application Data.
- Telecom Customer Segmentation
- Air Passangers Forecasting
- Predicting Loan Default
- Time Series Forecasting

COURSE OUTCOME

Ability to use advanced analytics techniques and tool to improve business performance across many functions by managing data with help of different tool like R. Working with various forms of structured and unstructured real time data sets in solving business problems with advanced statistical techniques & algorithms like machine learning

EXAM & CERTIFICATION

The certification is provided by Databyte Academy. Upon successful completion of the program, students will be conferred with **CERTIFIED DATA SCIENTIST**. In order to be "Certified" as part of the course, students need to complete the assignments and examination. Once all your assignments are submitted and evaluated, the certificate shall be awarded.

VISUALIZATION

- Creating Graphs
- Histograms & Density Plot
- Dot Plots – Bar Plots – Line Charts – Pie Charts – Boxplots – Scatterplots

BASIC STATISTICS (EXPLORATORY ANALYSIS)

- Univariate Analysis
- Bi-Variate Analysis (correlation, association etc)
- Descriptive Statistics(central tendency/variance)
- Frequency Tables /Summarization
- Exploratory Analysis
- Probability distributions
- Sampling – Central Limit Theorem
- Inferential statistics – Hypothesis testing
- Statistical tests (t/z-test, ANOVA, chi-square etc)

DATA PREP & REDUCTION TECHNIQUES

- Data Audit report creation and understanding
- Need for data preparation
- Binning, Dummy and Derived variable creation
- Standardization, Normalization
- Outlier treatment
- Missing values treatment (MI, clustering, regression based)
- Dimension reduction - Factor Analysis – PCA

CUSTOMER SEGMENTATION

- Basics of clustering
- Heuristic segmentation (RFM, Life stage, value based etc..)
- Cluster analysis (K-means and Hierarchical)
- Objective & subjective segmentation
- Decision Trees (CHAID/CART/C5.0)
- Cluster evaluation and profiling
- Interpretation and application

REGRESSION MODELLING

- Basics of regression analysis
- Approach: Model Estimation, OLS, MLE & Error Function for finding parameters, Assumptions verification (Linearity, Normality, multicollinearity, outliers etc.)
- Linear regression Model fitting
- Logistic regression Model Fitting
- Measures of goodness of fit (R^2 , Adj R^2 , Concordance, Gini, KS, Lift etc)
- Model Diagnostics – Residual Analysis– Decile Analysis – ROC Curves etc..
- Interpretation of results

TIME SERIES FORECASTING

- Time Series Introduction / Regression on Time, Time Series components
- Modelling Seasonality as Deviation
- Basic methods (pattern & pattern less)
- Averages (MA, WMA, CMA etc)
- Smoothing Techniques (Exponential)
- Advanced Methods (ARIMA etc)

INTRODUCTION TO MACHINE LEARNING

- Major Classes of Learning Algorithms -Supervised vs Unsupervised Learning
- Different Phases of Predictive Modelling
- Concept of Overfitting and Under fitting (Bias-Variance Trade off) & Performance Metrics
- Types of Cross validation(Train & Test, Bootstrapping, K-Fold validation etc)
- Cost & optimization functions

MACHINE LEARNING IN PRACTICE

- Ensemble Learning (Random Forest, Bagging & boosting)
- KNN, Naïve Bayes
- Introduction to Artificial Neural Networks(ANN)
- Introduction to Support Vector Machines(SVM)
- Introduction to Text Mining & NLP