



# Data Science



## **Course Overview:**

dridhOn's data science online course will make you gain knowledge in R programming, data manipulation with pandas, scikit learning, data modeling, data visualization with matplotlib, and so on. This data scientist training makes you solve complex real-time projects with all domains of data science, complete hands-on experience, The data science certificate offered by dridhon is approved in all large organizations in the world. These data science courses make your career much more beneficial.

## **Training Features:**

- 48 hours of blended learning
- 48 hours of Online self-paced learning 16 hours of instructor-led training
- Five lesson-end knowledge checks
- 1 real-life course-end project 20+ assisted practices on all modules
- Industry-recognized course completion certificate

## **Delivery Mode:**

- Online Live Virtual Instructor Led Training

## **Target Audience:**

- Business analyst and Business Intelligence
- Data analyst, data architects, and Software Engineering professionals who wish to develop their career
- Fresh graduates who interested to build their career in this domain can take up this online data science course

## **Key Learning Outcomes:**

This data science courses upskills you in ,

- Introduction to Data Science and its importance
- Understanding data types in programming languages
- Data operation and its manipulation
- Data modelling with KNN , decision tree
- Machine learning algorithms
- Data visualization and communication
- Data acquisition

## **Certification Details:**

- Complete at least 85 percent of the course or attend one complete batch
- Successful completion and evaluation of the project

## 1. Getting Started With Data Science And Recommender Systems

- Data Science Overview
- Reasons to use Data Science
- Project Lifecycle
- Data Acquirement
- Evaluation of Input Data
- Transforming Data
- Statistical and analytical methods to work with data
- Machine Learning basics
- Introduction to Recommender systems
- Apache Mahout Overview

## 2. Reasons To Use, Project Lifecycle

- What is Data Science?
- What Kind of Problems can you solve?
- Data Science Project Life Cycle
- Data Science-Basic Principles
- Data Acquisition
- Data Collection
- Understanding Data- Attributes in a Data, Different types of Variables
- Build the Variable type Hierarchy
- Two Dimensional Problem
- Co-relation b/w the Variables- explain using Paint Tool
- Outliers, Outlier Treatment
- Boxplot, How to Draw a Boxplot

## 3. Acquiring Data

- Discussion on Boxplot- also Explain
- Example to understand variable Distributions
- What is Percentile? – Example using Rstudio tool
- How do we identify outliers?
- How do we handle outliers?
- Outlier Treatment: Using Capping/Flooring General Method
- Distribution- What is Normal Distribution
- Why Normal Distribution is so popular
- Uniform Distribution
- Skewed Distribution
- Transformation

## 4. Machine Learning In Data Science

- Discussion about Box plot and Outlier
- Goal: Increase Profits of a Store
- Areas of increasing the efficiency
- Data Request
- Business Problem: To maximize shop Profits
- What are Interlinked variables
- What is Strategy
- Interaction b/w the Variables
- Univariate analysis
- Multivariate analysis
- Bivariate analysis
- Relation b/w Variables
- Standardize Variables
- What is Hypothesis?
- Interpret the Correlation
- Negative Correlation
- Machine Learning

## 5. Statistical And Analytical Methods Dealing With Data, Implementation Of Recommenders Using Apache Mahout And Transforming Data

- Correlation b/w Nominal Variables
- Contingency Table
- What is Expected Value?
- What is Mean?
- How Expected Value is differ from Mean
- Experiment – Controlled Experiment, Uncontrolled Experiment
- Degree of Freedom
- Dependency b/w Nominal Variable & Continuous Variable
- Linear Regression
- Extrapolation and Interpolation
- Univariate Analysis for Linear Regression
- Building Model for Linear Regression
- Pattern of Data means?
- Data Processing Operation
- What is sampling?
- Sampling Distribution
- Stratified Sampling Technique
- Disproportionate Sampling Technique
- Balanced Allocation-part of Disproportionate Sampling
- Systematic Sampling
- Cluster Sampling
- 2 angels of Data Science-Statistical Learning, Machine Learning

## 6. Testing And Assessment, Production Deployment And More

- Multi variable analysis
- linear regration
- Simple linear regration
- Hypothesis testing
- Speculation vs. claim(Query)
- Sample

## 7. Business Algorithms, Simple Approaches To Prediction, Building Model, Model Deployment

- Machine Learning
- Importance of Algorithms
- Supervised and Unsupervised Learning
- Various Algorithms on Business
- Simple approaches to Prediction
- Predict Algorithms
- Population data
- sampling
- Disproportionate Sampling
- Steps in Model Building
- Sample the data
- What is K?
- Training Data
- Test Data
- Validation data
- Model Building
- Find the accuracy
- Rules
- Iteration
- Deploy the model
- Linear regression

## 8. Getting Started With Segmentation Of Prediction And Analysis

- Clustering
- Cluster and Clustering with Example
- Data Points, Grouping Data Points
- Manual Profiling
- Horizontal & Vertical Slicing
- Clustering Algorithm
- Criteria for take into Consideration before doing Clustering
- Graphical Example
- Clustering & Classification: Exclusive Clustering, Overlapping Clustering, Hierarchy Clustering
- Simple Approaches to Prediction
- Different types of Distances: 1. Manhattan, 2. Euclidean, 3. Cosine Similarity
- Clustering Algorithm in Mahout
- Probabilistic Clustering
- Pattern Learning
- Nearest Neighbor Prediction
- Nearest Neighbor Analysis

## 9. Integration Of R And Hadoop

- R introduction
- How R is typically used
- Features of R
- Introduction to Big data
- R+Hadoop
- Ways to connect with R and Hadoop
- Products
- Case Study
- Architecture
- Steps for Installing RIMPALA
- How to create IMPALA packages