

Introduces Online Certificate Courses

Data Analysis With R Specialization



Preamble

With the increase in scale of production, the interaction of customers and business transactions are also largely escalating. This is eventually leading to large-scale generation of raw data. Hence, the need for data analysis becomes predominant for understanding the customer insights and better operability. It also helps in evaluating cost-benefit trade-offs through data through. Data-driven processes require data collection, data analysis, interpretation and visualization. Data analysis through open source programming tools like R assists the programmers and data managers to adopt best practices for revealing hidden patterns, unidentified correlations, evolving market trends, and customer behavior predictions.

This course is designed to impart the skills needed for analyzing big data and finding meaningful insights for decision making. The online program will consist of modules designed to enhance one's career prospects by expanding his knowledge in the area of analytics using R. Courses are designed specifically for adapting with the busy schedules of the professionals. The integration of theoretical learning and hands-on experimental learning is prominent part that compliments and substantiates the learning with practical orientation.

Duration of the Course: 3 months

Mode of Operation: Online

Number of Lectures of 1 hour duration: 24 L (12*2)

Number of Practical Sessions of 1 – 2 hours duration: 24 hours session (12*2)

To be covered in each Lecture

Every module will be covered in 2hrs

- | | |
|-----------|---|
| L1 | : Introduction to Data Analytics, Data Analytics types, Tools for Data Analytics, Types of data, Data Measurement Levels, Statistical tests for analyzing different data types, Introduction to R, Objects, Function, Conditional Statement, Loop, Scripts. |
| L2 | : Data Structures in R: Vector, Matrix, Array, Factor, List, Data Frame. |
| L3 | : String Handling, String Manipulation, Regular Expression, Pattern Matching. |

L4	: Working with Files, Import and export using csv and Excel, Import and export of Binary, XML and JSON file
L5	: Data Pre-processing using R: Packages, Missing data handling, Correlation, Data Transformation: Packages in R for data transformation, Scaling, Normalization, Yeo-Johnson Transform, PCA
L6	: Data Visualization: Packages available in R for visualization, Box Plot, QQ Plot, Scatter Plot, Pie Chart, Histogram, Plot Setting, Graph Data Analysis.
L7	: Supervised Learning Models: Classification Packages in R, Classification Techniques
L8	: Statistical parameters to evaluate performance of supervised learning models, Confusion Matrix, Accuracy, Sensitivity, Precision, Recall, ROC Curve, Plotting and analyzing model accuracy through AUC.
L9	: Regression Analysis: Linear & Multiple Regression, Logistic Regression, Regression Evaluation, Over fitting, Time Series Modelling: Introduction to Time series data, Moving Averages, Exponential Smoothing, ARIMA Model, Case Study.
L10	: Unsupervised Learning Models: Packages for clustering, Clustering Techniques and its Evaluation, Cluster goodness: Silhouette Method, Pseudo F-Statistic.
L11	: Hypothesis Testing, Analysis of Variance (ANOVA) – One way, Introduction to Correlation Analysis & its usage, Chi Square Test
L12	: Introduction to Association Analysis, Apriori Algorithm, Association rules.

To be covered in each Practical Session

P1	: Installation & Hands-on of R, Control Structure, Conditional Statements, Functions & Packages, Array & Matrix
P2	: Lists & Dictionaries, Data Frame, String handling and manipulation
P3	: Data Loading, Reading and writing data to different file format,
P4	: Data Cleaning, Data Preprocessing
P5	: Central Tendency Measures, Inferential & Descriptive Statistics
P6	: Data Transformation & Normalization
P7	: Data Visualization on varied case studies
P8	: Graph based data analysis
P9	: Implementation and Evaluation of Supervised Learning Models on Real Life Problems.
P10	: Implementation and Evaluation of Unsupervised Learning Models on Real Life Problems.
P11	: Hypothesis Testing, T-test, Z-test, ANOVA test, Chi square test
P12	: Market Basket Analysis

Pre-requisite, if any: Basic Computer science knowledge

Schedule of the Classes: (Saturday/ Sunday – 2 hours on each day)

Nature of the Course: Broad based

Name of the Faculty Coordinators: Dr. Megha Rathie & Dr. Neetu Sardana

Name(s) of the Faculty to be involved in conduction of the Course: Dr.MeghaRathi&Dr. Neetu Sardana

Target Participants: Professionals working in industry/Job seekers/Current Senior students who wish to upgrade their professional skills

Minimum Qualifications for participants: Preferably Undergraduate in any specialization also pursuing Undergraduate degree will also be considered.

IPR Rights of the Content: JIIT

Mode of evaluation of the participants after every 6 - 12 Lecture Sessions: MCQ and Subjective

Mode of evaluation of the participants after 3 – 6 Practical Sessions: Case Study based Evaluation

For course related query please mail to:

Dr.MeghaRathi : megha.rathi@jiit.ac.in Dr. Neetu Sardana : neetu.sardana@jiit.ac.in

For course registration, please click the link:

<https://forms.gle/AkVidsjMp5ArfwNj7>