SYLLABUS

UNIT – I

Introduction to Data Science and Overview of R

Data Science Process: Roles in a data science project, Stages in a data science project, Setting expectations. Basic Features of R, R installation, Basic Data Types: Numeric, Integer, Complex, Logical, Character. Data Structures: Vectors, Matrix, Lists, Indexing, Named Values, Factors. Subsetting R Objects: Sub setting a Vector, Matrix, Lists, Partial Matching, Removing NA Values. Control Structures: if-else, for Loop, while Loop, next, break. Functions: Named Arguments , Default Parameters, Return Values.

UNIT – II

Loading, Exploring and Managing Data

Working with data from files: Reading and Writing Data, Reading Data Files with read. table (), Reading in Larger Datasets with read.table. Working with relational databases. Data manipulation packages: dplyr, data.table, reshape2, tidyr, lubridate.

UNIT – III

Modelling Methods-I: Choosing and evaluating Models

Mapping problems to machine learning tasks: Classification problems, Scoring problems, Grouping: working without known targets, Problem-to-method mapping, Evaluating models: Over fitting, Measures of model performance, Evaluating classification models, Evaluating scoring models, Evaluating probability model.

UNIT – IV

Modelling Methods-II: Linear and logistic regression

Using linear regression: Understanding linear regression, Building a linear regression model, making predictions.
Using logistic regression: Understanding logistic regression, Building a logistic regression model, making predictions.

UNIT – V

Data visualization with R

Introduction to ggplot2: A worked example, Placing the data and mapping options, Graphs as objects, Univariate Graphs: Categorical, Quantitative.

Bivariate Graphs- Categorical vs. Categorical, Quantitative vs Quantitative, Categorical vs. Quantitative, Multivariate Graphs : Grouping, Faceting.