



SRI VENKATESWARA UNIVERSITY

I Year Syllabus for 4-Year UG Honours in B.Sc. (Data Mining) as Major in consonance with Curriculum framework w.e.f. AY 2025-26

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
I	I	1	Computer Fundamentals and Office Automation	3	3
			Computer Fundamentals and Office Automation-Practical	2	1
		2	Core Python Programming	3	3
			Core Python Programming-Practical	2	1
	II	3	Fundamentals of Data Mining	3	3
			Fundamentals of Data Mining-Practical	2	1
		4	R Programming Essentials for Data Mining	3	3
			R Programming Essentials for Data Mining-Practical	2	1

SEMESTER-II		
COURSE 4: R PROGRAMMING ESSENTIALS FOR DATA MINING		
Theory	Credits: 3	3 hrs/week

Aim and Objectives of the Course:

Aim:

To provide students with foundational programming skills in **R**, focusing on its applications in **data mining**, including data pre processing, exploration, visualization, and analysis.

Objectives:

- To understand the R environment and basic programming constructs.
- To develop data manipulation, cleaning, and transformation skills.
- To apply statistical and data mining techniques using R.
- To enable the visualization and interpretation of data insights.
- To prepare students for data-centric roles using R in analytics and mining projects.

Learning Outcomes of the Course:

After completing this course, students will be able to:

- Describe the R programming environment and data types.
- Write R scripts and use functions for automating data tasks.
- Load, clean, and transform raw data into analysis-ready formats.
- Visualize complex datasets using R's plotting libraries.
- Apply basic statistical and data mining methods using R.
- Handle real-world data mining problems using case studies and projects.

SYLLABUS:

UNIT 1: Fundamentals of R Programming: Overview of R and its applications in data mining, Installing R and RStudio, Data types: vectors, matrices, lists, factors, data frames, Variables, operators, and expressions, Reading and writing data (CSV, TXT).

UNIT 2:

Programming Constructs and Functions: Conditional statements (if, else, switch), Loops (for, while, repeat), Writing user-defined functions, Scope and environment, Error handling and debugging.

UNIT 3: Data Handling and Manipulation : Data cleaning and preprocessing techniques, Handling missing and duplicate data, Using dplyr, tidyr, stringr, and lubridate, Merging, filtering, and sorting datasets, Introduction to data wrangling pipelines

UNIT 4: Data Visualization and Exploratory Data Analysis (EDA): Base R plotting functions, **Introduction to ggplot2:** grammar of graphics, Visualizing distributions, trends, relationships, Multi-panel plots, themes, labeling, and annotations, Visual EDA for data mining

UNIT 5: Statistical Methods and Data Mining Applications: Descriptive statistics, Hypothesis testing (t-test, chi-square, ANOVA), Correlation and simple linear regression, Clustering (k-means introduction), Case studies: customer segmentation, market basket analysis.

Text Books:

1. **"The Art of R Programming"** by Norman Matloff, No Starch Press
2. **"R for Data Science"** by Hadley Wickham & Garrett Golemund, O'Reilly Media
3. **"Hands-On Programming with R"** by Garrett Golemund, O'Reilly Media

Reference Books:

1. **"Advanced R"** by Hadley Wickham, CRC Press
2. **"Data Mining with R: Learning with Case Studies"** by Luis Torgo, CRC Press
3. **"Machine Learning with R"** by Brett Lantz, Packt Publishing
4. **"Practical Data Science with R"** by Nina Zumel & John Mount, Manning Publications

SEMESTER-II		
COURSE 4: R PROGRAMMING ESSENTIALS FOR DATA MINING		
Practical	Credits: 1	2 hrs/week

Lab / Practical / Experiments / Tutorials Syllabus:

1. Installing R/ R Studio, basic syntax, operators
2. Working with vectors, matrices, lists, data frames
3. Writing custom functions and control structures
4. Data import/export exercises (CSV, Excel, etc.)
5. Data cleaning and transformation using dplyr.
6. Visualizing datasets with ggplot2
7. Descriptive and inferential statistics using R
8. Mini project: Real-life dataset analysis (e.g., COVID-19, crime, health data, etc.)

Recommended Co-Curricular Activities:

1. Participation in **data mining hackathons** using R
2. Workshops on **Data Visualization and Statistical Modeling**
3. Guest lectures from industry experts in **analytics/data science**
4. Contribution to R-based open-source projects on GitHub
5. Publishing articles or mini-projects on platforms like **Medium, R-bloggers**
6. Completing **online certifications** from platforms like Coursera (e.g., "*Data Science with R*")
7. Organizing a "**Data Challenge Day**" using R for solving real-world problems

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Mining Honours
I year II Semester
Course 4: R PROGRAMMING ESSENTIALS FOR DATA MINING

MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 3 marks.

(5X3=15)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer any **FIVE** of the following. Each Question Carries 12 marks.

(5X12=60)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V