



SRI VENKATESWARA UNIVERSITY

**4-Year UG Honours in B.Sc. (Data Mining) as Major in consonance with
Curriculum framework (W.E.F. AY-2026-27)**

COURSE STRUCTURE (for Semester III)

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
II	III	5	Data Base Management Systems	3	3
			Data Base Management Systems-Practical	2	1
		6	Data Structures & Algorithms using python	3	3
			Data Structures & Algorithms using Python-Practical	2	1
		7	Data Mining Techniques using R	3	3
			Data Mining Techniques using R -Practical	2	1



SRI VENKATESWARA UNIVERSITY

II Year Syllabus for 4-Year UG Honours in B.Sc. (Data Mining) as Major in consonance with the Curriculum framework w.e.f. AY 2026-27

II Year III Semester

Sl. No.	Category	Course No.	Course Name	No. of Hours		Total No. of Hours	No. of Credits		Total No. of Credits
				Theory	Practical		Theory	Practical	
1	Major-Core	5	Database Management Systems	3	2	5	3	1	4
2	Major-Core	6	Data Structures & Algorithms	3	2	5	3	1	4
3	Major-Core	7	Data Mining Techniques using R	3	2	5	3	1	4
4	Minor	1		3	2	5	3	1	4
5	AECC	1	English	4	0	4	3	0	3
6	AECC	1	MIL (Telugu/Hindi/Sanskrit)	4	0	4	3	0	3
7	MDC	2	-	2	0	2	2	0	2
8	SEC	3	SEC: Design Thinking/Problem solving/ others	2	0	2	2	0	2
				24	8	32	22	4	26

**SEMESTER-
III**

**COURSE 7: DATA MINING TECHNIQUES USING
R**

Theory

Credits: 3

3 hrs/week

Course Objectives:

1. To introduce the fundamentals of data mining concepts, techniques, and applications.
2. To develop skills in preprocessing and preparing data for mining.
3. To explore data mining evaluation methods for better decision-making.
4. To gain practical exposure by applying R programming to real-world datasets.
5. To encourage analytical thinking and problem-solving using data mining approaches.

Course Outcome:

By the end of the course, learners will be able to:

1. Explain the concepts, need, and applications of data mining.
2. Perform data cleaning, transformation, and visualization using R.
3. Implement unsupervised learning methods like clustering and association rule mining.
4. Use R programming to analyze real-world datasets and generate insights.

Syllabus:

UNIT 1:

Data Mining: Definition, importance, and applications, Knowledge Discovery in Databases (KDD) process, Data types, data sources, and data mining functionalities, Data mining challenges and issues, R programming basics for data mining.

UNIT 2:

Data cleaning - handling missing values, noise, outliers, Data transformation (normalization, discretization, aggregation), Feature selection and dimensionality reduction basics, Data visualization techniques in R - histograms, scatter plots, boxplots, heatmaps, Exploratory Data Analysis (EDA)

UNIT 3:

Basics of classification and prediction, Decision Trees – CART and C4.5 in R, Naïve Bayes classifier in R, k-Nearest Neighbors (k-NN) in R, Logistic Regression basics and implementation in R, Model evaluation metrics (confusion matrix, accuracy, precision, recall, F1-score)

UNIT 4:

Introduction to clustering methods, k-Means clustering in R, Hierarchical clustering in R, Density-based clustering basics, Association rule mining (Apriori algorithm) in R, Market Basket Analysis

UNIT 5:

Model evaluation and validation (cross-validation, hold-out), Overfitting and underfitting in data mining, Introduction to ensemble methods (Random Forest basics in R), Applications of data mining in healthcare, business, finance, and social media, Case studies using R classification + clustering + association rules.

ACTIVITIES:

1. Explore datasets (CSV/Excel) and summarize features using R.
2. Perform preprocessing on a dataset (remove null values, normalize data) in R

3. Visualize dataset attributes using R plots and interpret findings.
4. Compare performance of k-NN and Naïve Bayes on a chosen dataset.
5. Apply k-Means clustering on customer dataset and interpret clusters.
6. Perform cross-validation on a classification model in R.
7. Develop a mini-project applying at least two data mining techniques on a real dataset. **Prescribed Textbooks:**

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar – *Introduction to Data Mining*, Pearson.
2. Brett Lantz – *Machine Learning with R*, Packt Publishing.
3. K.P. Soman, Shyam Diwakar, V. Ajay – *Insight into Data Mining: Theory and Practice*, Prentice-Hall of India.

Reference Books:

1. Ian H. Witten, Eibe Frank, Mark Hall – *Data Mining: Practical Machine Learning Tools and Techniques*.
2. Galit Shmueli, Nitin R. Patel, Peter C. Bruce – *Data Mining for Business Intelligence*.
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman – *The Elements of Statistical Learning*.
4. Tanuja Sonker – *Data Mining Using R*, Lambert Academic Publishing.
5. Charu C. Aggarwal – *Data Mining: The Textbook*, Springer.

SEMESTER-III

COURSE 7: DATA MINING TECHNIQUES USING R

Practical

Credits: 1

2 hrs/week

List of experiments:

1. Load a dataset (CSV/Excel) into R and display summary statistics (mean, median, mode, standard deviation).
2. Handle missing values in a dataset using removal and imputation techniques in R.
3. Perform data normalization (Min-Max, Z-score) on numeric attributes using R functions.
4. Visualize dataset features using histograms, scatter plots, and boxplots in R.
5. Implement a Decision Tree classifier (using rpart or C50) on the Iris dataset and visualize the tree.
6. Apply Naïve Bayes classification (using e1071) on a dataset and evaluate with a confusion matrix.
7. Implement k-Nearest Neighbors (k-NN) on a dataset and compare accuracy with different k values.
8. Perform Hierarchical Clustering on a dataset and draw the dendrogram.
9. Use Apriori algorithm in R (arules package) to generate association rules from a transactional dataset.

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Mining Honours
II year III Semester
COURSE 7: Data Mining Techniques using R

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 **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