



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 3: FUNDAMENTALS OF DATA MINING

Theory

Credits: 3

3 hrs/week

Aim and Objectives of the Course:

Aim:

To provide fundamental knowledge and hands-on exposure to data mining techniques, their applications, and tools used for discovering patterns and insights from large datasets.

Objectives:

- To understand the principles of data mining and its real-world applications.
- To learn key data mining tasks: classification, clustering, association rules, and prediction.
- To explore algorithms such as ID3, K-Means, Apriori, and others.
- To provide knowledge on pre processing and data preparation for mining.
- To implement basic data mining operations using tools like **WEKA, R, or Python**.

Learning Outcomes of the Course:

Upon completion of the course, students will be able to:

- Understand data mining concepts and key functionalities.
- Apply data preprocessing techniques on raw datasets.
- Use classification, clustering, and association rule mining techniques effectively.
- Analyze and interpret patterns and trends from datasets.
- Implement and evaluate data mining algorithms using appropriate tools.
- Solve real-world problems using learned data mining methods.

UNIT 1: Introduction to Data Mining : Definition, Scope and Applications of Data Mining, Data Mining vs Machine Learning, KDD Process: Knowledge Discovery in Databases, Types of Data: Structured, Semi-structured, Unstructured, Major Issues in Data Mining.

UNIT 2: Data Preprocessing: Data Cleaning, Integration, Transformation, Data Reduction and Discretization, Feature Selection and Extraction, Handling Noisy and Missing Data, Normalization and Standardization Techniques.

UNIT 3: Classification and Prediction: Classification Concepts and Issues, Decision Tree (ID3, C4.5), Naive Bayes Classifier, k-Nearest Neighbor (k-NN), Model Evaluation and Accuracy Metrics, Overfitting and Underfitting.

UNIT 4: Clustering Techniques: Clustering vs Classification, Partitioning Methods: K-Means, K-Medoids, Hierarchical Clustering: Agglomerative and Divisive, Cluster Evaluation and Validity, Applications of Clustering in Real Life.

UNIT 5: Association Rule Mining: Basics of Association Rules, Apriori Algorithm: Support, Confidence, Lift, FP- Growth Algorithm, Rule Generation and Evaluation, Applications: Market Basket Analysis, Recommender Systems

Text Books:

1. **"Data Mining: Concepts and Techniques"**–Jiawei Han, Micheline Kamber, Jian Pei (Morgan Kaufmann)
2. **"Introduction to Data Mining"**–Pang Ning Tan, Michael Steinbach, Vipin Kumar (Pearson)
3. **"Fundamentals of Data Mining"**–Bhavani Thuraisingham (CRC Press)

Reference Books:

1. **"Data Mining Techniques"**– Arun K. Pujari (Universities Press)
2. **"Data Warehousing and Data Mining"**– M.A. Parthasarathy (Vikas Publishing)
3. **"Data Mining for Business Intelligence"**–Galit Shmueli, Nitin R. Patel, Peter C. Bruce (Wiley)
4. **"Machine Learning"**–Tom M. Mitchell (McGraw-Hill) Web

References for Case Studies:

1. <https://kaggle.com>–Real-world datasets and projects using data mining techniques
2. <https://www.r-bloggers.com>–Case studies and implementations using R for data mining
3. <https://archive.ics.uci.edu/ml/>–UCI Machine Learning Repository (benchmark datasets.)

SEMESTER-II
COURSE 3: FUNDAMENTALS OF DATA MINING

Practical

Credits: 1

2 hrs/week

Lab/Practical/Experiments/ Tutorials Syllabus:

1. Install and explore GUI-based and scripting-based tools such as **WEKA, Orange, R, and Python (scikit-learn)**. Understand data input formats and interface navigation.
2. Perform data cleaning, normalization, transformation, and missing value handling Using tools. Learn how to prepare data for mining tasks.
3. Apply **dimensionality reduction** and **feature engineering** techniques using WEKA or Python. Use filters and wrappers for feature selection.
4. Build and evaluate classification models using decision trees. Visualize decision paths and analyze classification accuracy.
5. Implement and compare **Naive Bayes** and **k-NN** classifiers. Understand assumptions and accuracy metrics.
6. Use **confusion matrix, precision, recall, F1 score, and ROC curves** for evaluating classifiers. Perform **cross-validation** and **holdout testing**.
7. Apply **K-Means** and **Agglomerative Hierarchical Clustering** algorithms. Visualize cluster results and interpret patterns.
8. Generate and interpret association rules from transaction datasets. Understand **support, confidence, and lift**. Use WEKA or Python libraries (e.g., mlxtend).
9. Create dashboards using **Excel, Tableau, or Python (Plotly/ Seaborn/ Matplotlib)** to visualize mining outcomes (e.g., clusters, associations, classifications).

Recommended Co-Curricular Activities:

1. Organize **workshops on data mining tools** like WEKA, Python, or R
2. Participation in **data mining or analytics competitions** (e.g., Kaggle challenges)
3. Encourage **mini-projects using open datasets** (e.g., retail, education, health, etc.)
4. Attend or presentations **student tech fests, data summits, or seminars**
5. Guest lecturers from industry professionals working in **data analytics or AI**
6. Certification from platforms like **Coursera, edX, or Data Camp** on data mining topics
7. Create a **department GitHub repository** for sharing code and case studies.

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Mining Honours
I year II Semester
Course 3: FUNDAMENTALS OF 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**