



SRI VENKATESWARA UNIVERSITY
II Year Syllabus for 4-Year UG Honours in BCA. (Data Science) as Major
in consonance with the Curriculum framework w.e.f. AY 2026-27

II Year IV 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	8	Python Programming	3	2	5	3	1	4
2	Major-Core	9	Document Oriented Database	3	2	5	3	1	4
3	Major-Core	10	Data Mining	3	2	5	3	1	4
4	Minor	2		4	0	4	4	0	4
5	MDC	3	-	2	0	2	2	0	2
6	SEC	4	SEC: Design Thinking/Problem solving/ others	2	0	2	2	0	2
				17	6	23	17	3	20

Note:

The number of instructional hours shall be five (5) for courses involving problem-solving, and four (4) for courses comprising only theory.


(A. Ravi Prasad)
CHAIRPERSON
BOS (PASSING) COMPUTER SCIENCE
S.V. UNIVERSITY
TIRUPATI



SRI VENKATESWARA UNIVERSITY

II Year Syllabus for 4-Year UG Honours in BCA Data Science as Major in consonance with Curriculum framework w.e.f. AY 2026-27

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
II	IV	8	Python Programming	3	3
			Python Programming – Practical	2	1
		9	Document Oriented Database	3	3
			Document Oriented Database – Practical	2	1
		10	Data Mining	3	3
			Data Mining – Practical	2	1

(A. Ravi Prasad)
CHAIRPERSON
BOS (PASS) IN COMPUTER SCIENCE
S.V. UNIVERSITY
TIRUMALA

SEMESTER-IV
BCA Data Science Honours
II Year IV Semester
COURSE 10 : Data Mining

Theory

Credits: 3

3 hrs/week

Course Objectives

- To understand the concepts, architecture, and design principles of Data Warehousing systems.
- To introduce core data mining concepts, preprocessing techniques, and similarity measures.
- To develop knowledge of association rule mining and frequent pattern discovery algorithms.
- To understand classification techniques including decision trees, rule-based, Bayesian, and nearest neighbor classifiers.
- To explore clustering paradigms and algorithms for structured and categorical data.

Course Outcomes

At the End of the Course, The Students will be able to:

- Explain Data Warehouse architecture, schema design (Star, Snowflake), OLAP cube, and OLAP operations.
- Describe data mining concepts, preprocessing techniques, similarity measures, and real-world applications.
- Implement association rule mining algorithms such as Apriori and FP-Growth for discovering frequent patterns.
- Apply classification techniques including Decision Trees, Bayesian, and Nearest Neighbor classifiers.
- Analyze and apply clustering algorithms such as K-Means, DBSCAN, BIRCH, and categorical clustering methods.

Unit-1: Data Warehousing:

Introduction to Data Ware House, Differences between Database systems and Data Ware House, Data Ware House characteristics, Data Ware House Architecture and its components, Data Modeling, Schema Design, star and snow-Flake Schema, Fact Constellation, Fact Table, OLAP cube, OLAP Operations.

Unit-2: Data Mining:

What is Data Mining? Data Mining: Definitions, KDD vs Data Mining, Data Mining Tasks, Data Preprocessing- Data Cleaning, Missing Data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binarization, Data Transformation; Measures of similarity and Dissimilarity-Basics. Issues and Challenges in DM, DM Applications- Case Studies.

Unit-3: Association Analysis:

Association Rules: What is an Association Rule?, Methods to Discover Association Rules, Apriori Algorithm, Partition Algorithm, Pincer-Search Algorithm, Dynamic Itemset Counting Algorithms, FP-Tree Growth Algorithm, Generalized Association Rule, Association Rules with Item Constraints

Unit-4: Classification:

Definition, What is Decision Tree?, Tree Construction Principle, Best Split, Splitting Indices, Splitting Criteria, Decision Tree Construction Algorithms: CART, ID3, C4.5, Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbor Classifiers, Bayesian Classifiers.


(A. K. Ravi)
HEADMASTER
BOS (PRACTICE) COMPUTER SCIENCE
S.V. UNIVERSITY
THIRUPATHI

Unit-5: Clustering Techniques:

Clustering Paradigms, Partitioning Algorithms (K-Means), k-Medoid Algorithms, Hierarchical Clustering: DBSCAN, BIRCH, Categorical Clustering Algorithms: STIRR, ROCK, CACTUS.

Textbooks:

1. Data Mining: Concepts and Techniques – Jiawei Han, Micheline Kamber & Jian Pei, Morgan Kaufmann.
2. Data Warehousing Fundamentals – Paulraj Ponniah, Wiley.

Reference Books:

1. Introduction to Data Mining – Pang-Ning Tan, Michael Steinbach & Vipin Kumar, Pearson.
2. Building the Data Warehouse – William H Inmon, Wiley.
3. The Data Warehouse Toolkit – Ralph Kimball, Wiley.

Activities:

Outcome: Illustrate Data Warehouse architecture, schema design techniques, and OLAP operations for decision support systems.

Activity: Create a block diagram and schema design for a sample business scenario (e.g., Retail Sales System) showing:

- o Data Warehouse architecture components
- o Star Schema and Snowflake Schema design
- o Fact table and dimension tables
- o OLAP cube representation
- o OLAP operations (Roll-up, Drill-down, Slice, Dice, Pivot)

Evaluation Method:

Poster presentation or case study submission. Use a rubric to assess correctness of schema design, clarity of OLAP operations, architectural representation, and explanation on a 10-point scale.

Outcome: Apply data preprocessing techniques and explain core data mining tasks and challenges.

Activity: Prepare a preprocessing report on a sample dataset demonstrating:

- o Data cleaning and handling missing values
- o Dimensionality reduction or feature selection
- o Data transformation techniques
- o Calculation of similarity/dissimilarity measures
- o Discussion on data mining challenges and applications

Evaluation Method:

Lab report and viva voce. Use a rubric to assess correctness of preprocessing steps, understanding of similarity measures, clarity of explanation, and case study relevance on a 10-point scale.

Outcome: Implement and compare association rule mining algorithms for discovering frequent patterns.

Activity: Perform market basket analysis on a sample dataset demonstrating:

- o Generation of frequent itemsets using Apriori algorithm
- o Support and confidence calculation

- o FP-Tree construction (conceptual or practical)
- o Interpretation of association rules
- o Comparison of algorithms (Apriori vs FP-Growth)

Evaluation Method:

Practical demonstration or analytical report. Use a rubric to assess correctness of algorithm steps, rule generation accuracy, interpretation of results, and comparison analysis on a 10-point scale.

Outcome: Apply classification algorithms and evaluate classifier performance.

Activity: Develop a classification model on a dataset demonstrating:

- o Decision Tree construction (ID3/C4.5/CART principles)
- o Selection of best split criteria
- o Implementation of Rule-based or Bayesian classifier
- o Comparison of classifiers using performance metrics

Evaluation Method:

Model demonstration and oral explanation. Use a rubric to assess correct algorithm implementation, split criteria justification, classifier comparison, and clarity of explanation on a 10-point scale.

Outcome: Analyze and implement clustering algorithms for numerical and categorical data.

Activity: Perform clustering analysis on a dataset demonstrating:

- o K-Means or K-Medoid clustering
- o Hierarchical clustering or DBSCAN/BIRCH
- o Interpretation of cluster results
- o Discussion on clustering paradigms and comparison of techniques

Evaluation Method:

Mini project presentation. Use a rubric to assess algorithm implementation accuracy, cluster interpretation, comparison of clustering methods, and presentation clarity on a 10-point scale.

SEMESTER-IV

COURSE 10: Data Mining

Practical

Credits: 1

2 hrs/week

List of lab experiments:

1. Load datasets in WEKA and explore data formats (ARFF/CSV)
2. Perform data cleaning and handle missing values using filters
3. Apply normalization and discretization on numeric attributes
4. Reduce data using attribute selection and PCA
5. Summarize and visualize data using statistical tools and class-wise comparison in WEKA.
6. Generate association rules using the Apriori algorithm
7. Apply multilevel association rule mining using hierarchical attributes
8. Apply K-means clustering and interpret the cluster outputs.
9. Perform hierarchical clustering and visualize results using dendrograms.
10. Apply Expectation-Maximization (EM) clustering and analyze cluster summaries.
11. Build a decision tree classifier using J48 and evaluate its performance.
12. Perform Naive Bayes classification and compare with decision tree results.
13. Apply rule-based classification using PART or JRip algorithms.
14. Compare classifiers using confusion matrix, accuracy, and ROC curves.
15. Perform basic text preprocessing and clustering using TF-IDF and K-means.


(A. Ravi Prasad)
BOS (PASS) IN COMPUTER SCIENCE
S.V. UNIVERSITY
TIRUPATI

BCA Data Science Honours
II Year IV Semester
COURSE 10 : Data Mining

Time : 3 Hrs

Max Marks : 75

SECTION - A

Answer any FIVE of the following

5 X 3= 15 Marks

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

SECTION-B

Answer FIVE of the following.

5 X 12=60 Marks

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


(A. Ravi Prasad)
CHAIRPERSON
BOS (PASSING COMPUTER SCIENCE)
S.V. UNIVERSITY
TIRUPATI