SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR
(AUTONOMOUS)
B.Tech.III Year I Semester RegularExaminations December-2025
DATA WAREHOUSING & DATA MINING
(Computer Science & Engineering)

Time: 3 Hours   **Max. Marks: 70**

PART-A
(Answer all the Questions 10 x 2 = 20 Marks)

1	a	Define Data Warehouse.	CO1	L1	2M
	b	List the phases of data warehouse development	CO1	L1	2M
	c	Name any two visualization tools.	CO2	L1	2M
	d	List any three data reduction techniques.	CO2	L1	2M
	e	Mention the need for multi-level mining.	CO3	L1	2M
	f	What is predictive accuracy?	CO3	L1	2M
	g	What are lazy learners?	CO4	L1	2M
	h	Mention two uses of outlier detection	CO4	L1	2M
	i	State any two challenges in Auto Imports dataset.	CO5	L1	2M
	j	List two clustering evaluation measures	CO5	L1	2M

PART-B
(Answer all Five Units 5 x 10 = 50 Marks)

UNIT-I

2	a	Illustrate a simple architecture of a data warehouse and explain its components.	CO1	L3	5M
	b	Compare MOLAP, ROLAP, and HOLAP based on performance and storage.	CO1	L2	5M

OR

3	a	Explain hierarchical levels in dimensions with examples.	CO1	L4	5M
	b	Explain the characteristics of OLAP systems.	CO1	L2	5M

UNIT-II

4	a	Explain the importance of data preprocessing.	CO2	L4	5M
	b	Describe data reduction techniques.	CO2	L2	5M

OR

5	a	Differentiate similarity/dissimilarity metrics for clustering tasks.	CO2	L4	5M
	b	Analyze the challenges and issues in data mining system design.	CO2	L2	5M

UNIT-III

6	a	Explain constraint-based pattern generation.	CO3	L3	5M
	b	Compare strong and weak constraints.	CO3	L2	5M

OR

7	a	Explain the process of associative classification with examples.	CO3	L2	6M
	b	Discuss rule generation and rule selection.	CO3	L3	4M

UNIT-IV

8	a	Discuss the strengths and limitations of naïve Bayes classification.	CO4	L3	6M
	b	Explain the working of SVM with a neat diagram.	CO4	L2	4M

OR

9	a	Discuss data distribution–based outlier detection.	CO4	L4	5M
	b	Explain k-means clustering with an example.	CO4	L3	5M

UNIT-V

10	a	Compare the classes of Breast Cancer dataset.	CO5	L3	5M
	b	Discuss preprocessing and filters in WEKA.	CO5	L2	5M

OR

11	a	Compare different classifiers for high-dimensional datasets.	CO5	L4	5M
	b	Analyze Apriori results using support and confidence thresholds.	CO5	L2	5M

*** END ***