

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY-GURUJADA VIZINAGARAM
III B. Tech I Semester Regular Examinations November -2025
DATA WAREHOUSING AND DATA MINING
(CSE)

Time: 3 hours

Max. Marks: 70

The Question paper consists of Part A & Part B.

Part A is compulsory, Answer all questions. Part B Answers any one each unit.

1		PART-A	(20Marks)
	a)	Differentiate between OLTP and OLAP.	[2]
	b)	List two major issues in Data Mining.	[2]
	c)	Define Data Transformation and provide an example.	[2]
	d)	What is a dissimilarity measure ?	[2]
	e)	Give an example of a classification problem in the real world.	[2]
	f)	What is the purpose of a Holdout method in classification model evaluation?	[2]
	g)	Define Support and Confidence in the context of association rules.	[2]
	h)	Briefly explain the concept of a Closed Itemset .	[2]
	i)	State the primary objective of Cluster Analysis .	[2]
	j)	What are the main strengths of the K-Means clustering algorithm?	[2]
		PART-B	(50Marks)
		Unit - I	
2	a)	Explain the architecture of a Data Warehouse with a neat diagram.	[5]
	b)	Elaborate on the different Data Mining Functionalities. What kinds of patterns can be mined?	[5]
		(OR)	
3	a)	Describe the different OLAP Operations in a Multidimensional Data Model using a suitable example.	[5]
	b)	Detail the Knowledge Discovery Process from Databases, illustrating the various steps involved.	[5]
		Unit - II	
4	a)	Explain the various reasons <i>Why</i> we need to pre-process the data before a mining application.	[5]
	b)	Describe the techniques involved in Data Reduction .	[5]
		(OR)	
5	a)	Discuss Data Cleaning techniques.	[5]
	b)	Explain different measures for calculating Data Similarity and Dissimilarity between data objects.	[5]
		Unit - III	
6	a)	Explain the general approach to solving a classification problem.	[5]
	b)	Discuss Bayesian Classification and its underlying principle.	[5]
		(OR)	
7	a)	Explain the concepts of Accuracy and Error measures used for classification model evaluation.	[5]
	b)	What is a Decision Tree ? Describe the concept of Tree Pruning and why it is necessary.	[5]
		Unit - IV	
8	a)	Define Market Basket Analysis and explain the importance of Frequent Itemsets and Association Rules in this context.	[5]
	b)	Illustrate the steps of the Apriori Algorithm for finding Frequent Itemsets using candidate generation with an example.	[5]
		(OR)	

9	a)	Discuss the FP-Growth Algorithm as a pattern growth approach	[5]
	b)	Explain the procedure for Generating Association Rules from Frequent Itemsets.	[5]
		Unit - V	
10	a)	Explain the K-Means Partitioning Method for clustering.	[5]
	b)	Describe the difference between Agglomerative and Divisive Hierarchical Clustering methods.	[5]
		(OR)	
11	a)	Explain the concept of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm.	[5]
	b)	Discuss the different Types of Clusters that can be found in data sets.	[5]
