Reg. No.															

Question Paper Code	13918

M.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

Fifth Semester

M.Tech. - Computer Science and Engineering (5 Years Integrated)

20CJPC501 – DATA WAREHOUSING AND DATA MINING

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	Arranging the customers names in ascending order is an example of (a) Information processing (b) Process (c) data (d) none of the above	1	K1	CO1
2.	In the context of data warehousing, what does the acronym ETL stand for? (a) Extract, Transform, Load (b) Evaluate, Transform, Load (c) Extract, Transfer, Load (d) Evaluate, Transfer, Load	1	K1	CO1
3.	Data Visualization in mining cannot be done using (a) Photos (b) Graphs (c) Charts (d) Information Graphics	1	K1	CO2
4.	Which of the following techniques is often used for classification? (a) K-Means Clustering (b) Decision Trees (c) Apriori Algorithm (d) PCA	1	K1	CO2
5.	Which of the following is a common metric used to evaluate association rules? (a) Mean Squared Error (b) Support Variance (c) Correlation (d) Coefficient	1	K1	CO3
6.	What does "confidence" measure in the context of association rules? (a) The likelihood of an item being purchased alone (b) The strength of an association rule (c) The total number of transactions (d) The overall popularity of an item	1	K1	CO3
7.	How many terms are required for building a bayes model? (a) 1 (b) 2 (c) 3 (d) 4	1	K1	CO4
8.	What is the main purpose of tree pruning in a decision tree classifier? (a) Increase the size of the decision tree to fit all training data perfectly (b) Remove branches that reflect noise or outliers to improve generalization (c) Add more attributes to increase accuracy on the training dataset (d) Randomly delete nodes to reduce computation time	1	K1	CO4
9.	Which R function is used for K-Means clustering? (a) hclust() (b) kmeans() (c) cutree() (d) dist()	1	K1	CO5
10.	In clustering high-dimensional data, which technique is often used to reduce dimensionality before clustering? (a) Linear Regression (b) Principal Component Analysis (PCA) (c) K-Nearest Neighbors (KNN) (d) Decision Trees	1	K1	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

		Marks	K – Level	CO
11.	What is meant by star Schema?	2	K1	CO1
12.	Write any two characteristics of OLAP systems.	2	K2	CO1
13.	Identify data characterization related to data discrimination.	2	K2	CO2
14.	Compare and contrast different types of reductions.	2	K2	CO2
15.	Define Sequential pattern mining.	2	K1	CO3
16.	What is meant by quantitative association rule?	2	K1	CO3
17.	Apply the k-means concept: How many clusters will be formed if k = 3 and there are 15 data points?	2	K2	CO4
18.	List out the Bayes classification methods.	2	K1	CO4

19.	What is the purpose of the arules package in R?	2	K1	CO5
20.	Show how to count the number of NA values in a data frame column.	2	K2	CO5
21.	List the categories of clustering methods.	2	K1	CO6
22.	Define outlier. How will you determine outliers in the data?	2	K1	CO6

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a)	Explain mapping data warehouse with multiprocessor architecture with the concept of parallelism and data partitioning.	11	K2	CO1
		OR			
	b)	Explain the multidimensional database schemas in detail.	11	K2	CO1
24.	a)	Elaborate the various classifications of data mining systems with example.	11	K2	CO2
		OR			
	b)	Illustrate the Major issues in data warehousing and data mining.	11	K2	CO2
25.	a)	Write an algorithm for FP-Tree Construction and discuss how frequent itemsets are generated from FP-Tree.	11	K2	CO3
		OR			
	b)	Discuss in detail about constraint and correlation based association mining.	11	K2	CO3
26.	a)	Summarize all the rules used for classification in Rule Based Classifier with a suitable example.	11	K2	CO4
		OR			
	b)	Demonstrate the process of classification by Decision tree induction.	11	K2	CO4
27.	a)	Implement the conceptual R commands (e.g., using the lm() function) to build a Multiple Linear Regression model for predicting a continuous outcome variable (Y) using three predictor variables (X_1, X_2, X_3). Given the R output summary that shows the coefficients, interpret the meaning of the coefficient value for X_2 in the context of the model.	11	K3	CO5
		OR			
	b)	Develop a R script for performing Bayesian classification on a sample dataset.	11	K3	CO5
28.	a)	Construct a hierarchical clustering model for a sample dataset using R programming.	11	K3	CO6
		OR			
	b)	Apply the k mediods algorithm on the following points and form the 2 clusters. x1=(2,6), x2=(3,4), x3=(3,8), x4=(4,7), x5=(6,2), x6=(6,4), x7=(7,3) x8=(7,4), x9=(8,5), x10=(7,6).	11	K3	CO6