End Term (Odd) Semester Examination November 2025

Roll No:

Name of the Program: B. Tech CSE AI/ML — Semester: VII

Name of the Paper: Data Warehousing and Data Mining — Paper Code: TCS-722

Time: 3hr — Max Marks: 100

Note:

i. All questions are compulsory.
ii. Answer any two sub questions among a, b & c in each main questions
iii. Total marks in each main question are twenty.
iv. Each questions carries 20 marks

Q1	(20Marks)	
(a)	Define a **Data Warehouse**. Discuss its architecture, key features, and advantages in decision support systems.	CO1
(b)	Explain the **Three-Tier Architecture** of a Data Warehouse with a neat, labeled diagram. Discuss the role of each tier (bottom, middle, top).	CO1
(c)	Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. (a) Enumerate three classes of schemas that are popularly used for modeling data warehouses. (2Marks) (b) Draw a schema diagram for the above data warehouse using one of the schema classes listed in (a). (3Marks) (c) Starting with the base cuboid [day, doctor, patient], what specific OLAP operations should be per- formed in order to list the total fee collected by each doctor in 2014? (2Marks) (d) To obtain the same list, write an SQL query assuming the data is stored in a relational database with the schema fee (day, month, year, doctor, hospital, patient, count, charge). (3 Marks)	CO1
Q2	(20Marks)	CO2
(a)	Describe the **OLAP operations** — roll-up, drill-down, slice, dice, and pivot — with suitable examples and diagrams.	CO2
(b)	Describe in detail the **backup and recovery process** for a large-scale data warehouse. Explain different backup strategies (full, incremental, differential).	
(c)	Explain how **performance tuning and testing** contribute to maintaining the quality and efficiency of a data warehouse system.	CO2
Q3	(20Marks)	
(a)	Define **data mining**. Explain its **functionalities** such as classification, clustering, association rule mining, and prediction with examples.	CO3
(b)	Suppose that the data for analysis includes the attribute *age*. The *age* values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36. 40, 45, 46, 52, 70. Use *smoothing by bin means* to smooth the above data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data.	CO3
(c)	Explain the concept of **numerosity reduction**. Compare **parametric** and **non-parametric** methods with examples.	CO3

End Term (Odd) Semester Examination November 2025

Q4		
	(20Marks)	
(a)	Explain in detail the process of **data generalization** and **analytical characterization** with suitable diagrams.	CO4
(b)	A healthcare organization maintains patient data across multiple hospitals. • Explain how **attribute relevance analysis** and **class comparison** can help identify key health indicators. • Describe how **association rules** could be mined to find relations between symptoms and diseases.	CO4
(c)	Explain in detail the **Apriori algorithm** with an example dataset. • Include steps for generating frequent itemsets and rules. • Show how support and confidence are calculated	CO4
Q5		
	(20Marks)	CO5
(a)	Discuss the **issues** regarding classification and prediction, including data quality, overfitting, missing values, and model evaluation.	
(b)	Explain K-nearest neighbor estimation? Consider the following data: Classify weather a special tissue paper with (x1=3 and x2=7) is good or not on the basic of two attribute using k-nearest neighbor approach.	CO5

Acid Durability (X1)	Strength (X2)	Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

(c)	A financial institution wants to detect **fraudulent transactions** from millions of customer records. • Explain how **DBSCAN** or **OPTICS** can help detect outliers. • Suggest how **outlier analysis** can enhance fraud detection.	CO5