

TE Sem V - CMAN - R19 - C-Scheme Winter 2025
19/11/25

Time: 3 Hrs

Max. Marks: 80

- Note (1). Question No.1 is compulsory.
 (2). Out of the remaining attempt any three.
 (3). Assume and mention suitable data wherever required.

Q.1 Solve any four of the following:(5 marks each) (20)

- A. Define data exploration and explain various types of data attributes commonly used in data mining
 B. List out steps of the K- Medoid algorithm.
 C. Write a short note on types of multilevel association rule mining.
 D. Explain web mining in brief.
 E. Describe random sub sampling and k fold cross validation accuracy measures of classification.
 F. Differentiate ER Modelling vs Dimensional Modeling.

Q.2 A. Consider the minimum support count is 2 and minimum confidence is 60% .
 Apply Apriori algorithm to find frequent itemsets and strong association rules. (10)

Tid	Items Purchased
1	Book, Pen
2	Book, Notebook, Pen
3	Notebook, Pen
4	Book, Notebook
5	Pen, Pencil

Q.2 B. Explain different methods of data visualization. (10)

Q.3 A. Consider the data given below. Create an adjacency matrix and apply a complete link agglomerative clustering algorithm to cluster the given data set and draw the dendrogram. (10)

Object	Attribute 1 (X)	Attribute 2 (Y)
A	2	2
B	3	2
C	1	1
D	3	1
E	1.5	1.5

Q.3 B. Consider a data warehouse for Bank Transactions with dimensions:

- (a) Branch (b) Account Type and (c) Time, and a measure as the Total Transaction Amount. Construct a cube and demonstrate the OLAP operations: Roll-up, Drill-down, Slice, Dice, and Pivot. (10)

Q.4 A. What is dimensional modeling? Suppose that a data warehouse for Big-University consists of the following four dimensions: student, course, semester, and instructor, and two measures count and avg_grade. Draw information package diagram and snowflake schema diagram for this data warehouse. (10)

Q.4 B. The company wants to predict whether an employee will get a bonus based on Experience and Project Performance Create a classification model using a decision-tree algorithm to predict the bonus and derive the classification rules. (10)

Tid	Experience	Performance	Bonus
1	High	Good	Yes
2	Low	Good	No
3	Medium	Poor	No
4	High	Poor	Yes

Q.5 A. Explain Page Rank algorithm with example. (10)

Q.5. B. Explain what different data cleaning and data transformation tasks are required before applying data mining techniques on the data set. (10)

Q.6. A. Explain Frequent Pattern tree algorithm with an example. (10)

Q.6. B Explain the following in brief: (10)

- i) Techniques of data extraction.
- ii) Naive bayes classification.
