

T.E. CMPN SEM V C-Scheme June 2026

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.

10/06/26

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

(20)

- Explain components of Data warehouse architecture.
- What is Data cleaning? What are the ways to fill the missing values?
- Explain any one partitioning cluster method.
- Explain support and confidence measures with formulas.
- Describe cross validation and Bootstrap methods for classification accuracy evaluation.
- Write a short note on Page rank algorithm.

Q.2 A. Construct step by step FP tree and find frequent item sets for a given dataset. Consider Minimum support count as 2. (10)

Tid	Items
10	A,C,D
20	B,C,E
30	A,B,C,E
40	B,E

Q.2 B. What is Dimensional Modeling? Consider a College Data Warehouse with the following dimensions: Student, Subject, Department, Academic Year and Measures: Marks Obtained, Attendance Percentage. Draw Star and Snowflake Schema Diagram. (10)

Q.3 A. Consider the data given below. Create a distance matrix using Euclidean distance and apply a single link hierarchical clustering algorithm to cluster the given data set and draw the dendrogram. (10)

Object	Attribute 1 (X)	Attribute 2 (Y)
P	2	3
Q	3	3
R	6	5
S	7	6
T	5	4

- Q.3. B. i) Explain KDD process with diagram. (5)
 ii) Describe applications of data mining. (5)

Q.4 A. Explain the following: Descriptive data summarization, Cleaning, Integration & Transformation. (10)

Q.4 B. Consider the following dataset for predicting whether a student will Pass or Fail based on Study Hours and Attendance. (10)

Study Hours	Attendance	Result
High	Good	Pass
High	Poor	Pass
Medium	Good	Pass
Low	Good	Fail
Low	Poor	Fail
Medium	Poor	Pass
High	Good	Pass
Low	Good	Fail

Using Naive Bayes Classification algorithm, classify the following tuple:
Study Hours = Medium, Attendance = Good

Q.5 A. Explain Clever algorithm with example. (10)

Q.5 B. Consider a data warehouse for Library Data Warehouse with dimensions:

(a) Book Category (b) User Type and (c) Time, and a measure as the Number of Books Issued. Construct a cube and demonstrate the OLAP operations: Roll-up, Drill-down, Slice and Dice. (10)

Q.6. A. What is multilevel association rule mining? Explain its types with an example. (10)

B. Write short note on following: (10)

i) ETL.

ii) Decision tree classification.
