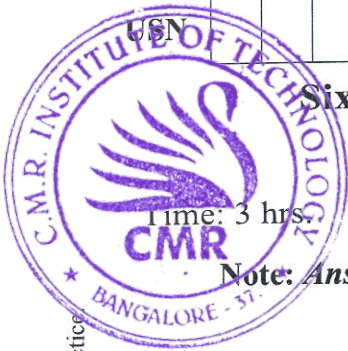


CBCS SCHEME

17CS651



Sixth Semester B.E. Degree Examination, Jan./Feb. 2023 Data Mining and Data Warehousing

Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Module-1

- a. Describe a 3 – tier data warehousing architecture. (08 Marks)
b. Compare OLTP and OLAP system. (06 Marks)
c. What is a Data Warehouse and what are its four key features? (06 Marks)

OR

- a. Explain with suitable examples the various OLAP operations in Multidimensional data model. (10 Marks)
b. Explain the following with examples : (10 Marks)
i) Snow flake schema ii) Fact constellation schema.

Module-2

- a. Describe ROLAP , MOLAP , HOLAP. (06 Marks)
b. What is Data Mining? With a neat diagram, explain KDD process in Data Mining. (08 Marks)
c. For the following vector X and Y , calculate the Cosine similarity where
 $X = \{3\ 2\ 0\ 5\ 0\ 0\ 0\ 2\ 0\ 0\}$ $Y = \{1\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 2\}$. (06 Marks)

OR

- a. Describe the various types of attributes and data sets. (10 Marks)
b. Define Data Pre – processing. Mention the steps involved in it. Explain any 2 steps in detail. (10 Marks)

Module-3

- a. Explain the Apriori Algorithm for frequent itemset generation with example. (10 Marks)
b. Generate Frequent itemset for the given data with support = 50%.

TID	100	200	300	400
Items	{1, 3, 4}	{2, 3, 5}	{1, 2, 3, 5}	{2, 5}

(10 Marks)

OR

- a. Consider the following transaction set : (10 Marks)
i) Construct an FP tree ii) Generate the list of frequent item set.
Order by their corresponding suffix.

TID	1	2	3	4	5	6	7	8	9	10
Items	{a, b}	{b, c, d}	{a, c, d, e}	{a, d, e}	{a, b, c}	{a, b, c, d}	{a}	{a, b, c}	{a, b, d}	{b, c, e}

- b. Briefly explain the candidate generation procedure using $F_{k-1} \times F_{k-1}$ merging strategy. (10 Marks)

Module-4

- a. Explain how decision tree induction algorithm works, with example. (10 Marks)
b. List and explain the different characteristics of decision tree induction. (10 Marks)

OR

- 8 a. Describe the nearest neighbour classification techniques. (10 Marks)
b. Write a note on Bayesian classifier. (10 Marks)

Module-5

- 9 a. What is Cluster Analysis? Describe the different types of clustering techniques with examples. (10 Marks)
b. Explain the following terms :
i) K – means clustering ii) Graph based clustering. (10 Marks)

OR

- 10 a. What are the approaches used for generating Agglomerative Hierarchical Clustering? (10 Marks)
b. Explain DB Scan algorithm with example. (10 Marks)
