



CBCS SCHEME

USN

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

17CS651

Sixth Semester B.E. Degree Examination, Jan./Feb. 2021 Data Mining and Data Warehousing

Time: 3 hrs.

Max. Marks: 100

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

Module-1

- 1 a. Define Data Warehouse. Explain with a neat diagram Data warehouse Architecture. (10 Marks)
- b. Compare OLTP and OLAP Systems. (06 Marks)
- c. What are the advantages of deploying data warehouse? (04 Marks)

OR

- 2 a. Explain the 3 different schemas of multi dimensional data models. (10 Marks)
- b. Discuss SLICE and DICE OLAP Operations. (04 Marks)
- c. What is Meta data in data warehousing? What are the contents of metadata repository? (06 Marks)

Module-2

- 3 a. Discuss ROLAP, MOLAP and HOLAP Server architecture. (09 Marks)
- b. Define Similarity and dissimilarity between the objects. Find COSINE, EUCLIDIAN and JACCARD similarity measures between the following two binary vectors.
 $X = (0, 1, 0, 1)$ and $Y = (1, 0, 1, 0)$. (11 Marks)

OR

- 4 a. What do you mean by Data Preprocessing? Explain any four tasks of Data Preprocessing in detail. (10 Marks)
- b. Describe the challenges that motivated the development of Data mining. (08 Marks)
- c. Mention the various types of attributes in a dataset. (02 Marks)

Module-3

- 5 a. Generate frequent item set and demonstrate Association rules for the following transaction data base with min. support = 40% and min. confidence = 60%. (12 Marks)

TID	Date	Item sets
T100	10/15/17	{K, A, D, B}
T200	10/15/17	{D, A, C, E, B}
T300	10/19/17	{C, A, B, E}
T400	10/22/17	{B, A, D}

- b. Describe Alternate methods for generating Frequent item sets. (08 Marks)

OR

- 6 a. Explain briefly FP – Growth algorithm. (10 Marks)
- b. Describe the various measures of evaluating Association Patterns. (10 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

Module-4

- 7 a. Apply decision tree algorithm to identify but split (information gain) for the following data based on the attribute outlook. (12 Marks)

ID	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
6	Rainy	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

- b. Explain the characteristics of decision tree induction algorithm. (08 Marks)

OR

- 8 a. Explain K – nearest neighbor classification algorithm with example. (10 Marks)
 b. Describe how Bayes theorem can be used for solving a classification problem. (10 Marks)

Module-5

- 9 a. What is Cluster Analysis? Explain the requirements of cluster analysis. (10 Marks)
 b. Explain briefly Agglomerative hierarchical clustering with example. (10 Marks)

OR

- 10 a. Explain DBSCAN algorithm with example. (08 Marks)
 b. Illustrate Grid – based clustering algorithm. How clusters are formed from Dense Grid cells. (12 Marks)
