

CBCS SCHEME

USN

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

15CS651

Sixth Semester B.E. Degree Examination, July/August 2022

Data Mining and Data Warehousing

Time: 3 hrs.

Max. Marks: 80

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

Module-1

- 1 a. Define Data-Warehouse and explain four key features of data-warehouse compare to other data-repository systems. (05 Marks)
- b. Explain the data-warehouse multi-architecture with neat diagram. (05 Marks)
- c. Define Operational Data-Store (ODS). With neat diagram explain the working of ODS design and implementation. (06 Marks)

OR

- 2 a. Define Extraction-Transformation Loading (ETL). Explain the processing steps of ETL. (05 Marks)
- b. List the different Data-model for Data-warehouse and explain any two with neat diagram. (06 Marks)
- c. Describe the different OLAP operations with an example. (Any two operations) (05 Marks)

Module-2

- 3 a. Explain efficient computation of datacube by creating the data cube for electronic sales that contains the following city, item, year and sales-in-dollars, with a neat diagram for following queries: i) Compute the sum of sales, grouping by city and item.
ii) Compute the sum of sales, grouping by city.
Consider 1D, 2D and 3D Cuboid. (06 Marks)
- b. Describe the difference between ROLAP and MOLAP. (05 Marks)
- c. Explain three kinds of data-warehouse applications. (05 Marks)

OR

- 4 a. Define Data-Mining. Explain the process of Knowledge Discovery in Database (KDD). (06 Marks)
- b. Explain different challenges that motivated the development of Data-mining technologies (any 5). (05 Marks)
- c. Describe different data-preprocessing approaches (any 5). (05 Marks)

Module-3

- 5 a. Develop the Apriori algorithm for generating frequent itemset. (08 Marks)
- b. Consider the transaction data-set:

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

Construct the FP tree by showing the trees separately after reading each transaction.

(08 Marks)

OR

- 6 a. Consider the following transaction data-set 'D' shows 9 transactions and list of items using Apriori Algorithm frequent - itemset minimum support for 2.

Tid	List of Items
T ₁	I ₁ , I ₂ , I ₅
T ₂	I ₂ , I ₄
T ₃	I ₂ , I ₃
T ₄	I ₁ , I ₂ , I ₄
T ₅	I ₁ , I ₃
T ₆	I ₂ , I ₃
T ₇	I ₁ , I ₃
T ₈	I ₁ , I ₂ , I ₃ , I ₅
T ₉	I ₁ , I ₂ , I ₃

- b. Write a short note on Simpson's paradox (04 Marks)
- c. Define association-analysis. Explain the association rule representation with an example. (04 Marks)

Module-4

- 7 a. Define classification. Explain the general approach for solving classification model. (08 Marks)
- b. Write an algorithm for Decision-tree induction. (08 Marks)

OR

- 8 a. List the important characteristics of Decision-tree induction. (06 Marks)
- b. Explain rule-based classifiers used for classification. (05 Marks)
- c. Write k-nearest neighbor classification algorithm. (05 Marks)

Module-5

- 9 a. What is Cluster Analysis? Explain the different types of clustering. (05 Marks)
- b. Explain K-mean clustering method and algorithm. (06 Marks)
- c. Explain Agglomerative Hierarchical clustering algorithm for computing distances between clusters. (05 Marks)

OR

- 10 a. How Density based methods are used for clustering (DBSCAN Algorithm). Explain with example. (05 Marks)
- b. Explain MST clustering with the help of algorithm. (05 Marks)
- c. Write a short note on:
- BIRCH
 - CURE
- (06 Marks)

* * * * *