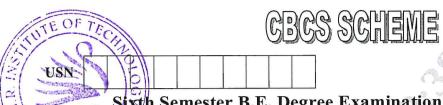
BANGALURE Time: 3 hrs.



15CS651

Sixth Semester B.E. Degree Examination, July/August 2022 **Data Mining and Data Warehousing**

Max. Marks: 80

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

Module-1

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

- Define Extraction-Transformation Loading (ETL). Explain the processing steps of ETL. (05 Marks)

(05 Marks)

- List the different Data-model for Data-warehouse and explain any two with neat diagram. (06 Marks)
- Describe the different OLAP operations with an example. (Any two operations) (05 Marks)

Module-2

- Explain efficient computation of datacube by creating the data cube for electronic sales that 3 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)

Describe the difference between ROLAP and MOLAP.

(05 Marks)

Explain three kinds of data-warehouse applications.

(05 Marks)

- 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 (05 Marks) (any 5).
 - Describe different data-preprocessing approaches (any 5).

(05 Marks)

Module-3

Develop the Apriori algorithm for generating frequent itemset.

(08 Marks)

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.

Det IIIII	op p
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

(08 Marks)

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

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:
 - i) BIRCH
 - ii) CURE

(06 Marks)

* * * * :