

CBCS SCHEME

USN

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

15CS651

Sixth Semester B.E. Degree Examination, Feb./Mar. 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. Explain multi-tier architecture of data warehouse. (09 Marks)
b. Write the comparisons between OLTP and OLAP systems. (07 Marks)

OR

- 2 a. Explain the schemas for multidimensional data models. (08 Marks)
b. Explain typical OLAP operations. (08 Marks)

Module-2

- 3 a. Discuss data mining tasks with examples. (10 Marks)
b. Explain OLAP servers. (06 Marks)

OR

- 4 a. Define Attribute. Explain types of attributes. (07 Marks)
b. For the following vectors X and Y. Calculate the Hamming, cosine, correlation, Jaccard distance measures.
X = (1, 1, 0, 1, 0, 1)
Y = (1, 1, 1, 0, 0, 1) (09 Marks)

Module-3

- 5 a. What is frequent Itemset Generation? Write an apriori algorithm to generate frequent itemset. (08 Marks)
b. Consider the following transaction data set 'D' shows 9 transactions and list of items-using Apriori algorithm to find frequent itemset min-support threshold 22%.

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

(08 Marks)

OR

- 6 a. Consider the transaction data set

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}

Constrict FP-tree by showing the trees separately after reading each transaction . ID.

(08 Marks)

- b. Write the alternative methods that have been developed to overcome the limitations of the Apriori algorithm. (08 Marks)

Module-4

- 7 a. Explain Hunt's algorithm for inducing decision trees. And Write the design issues associated with it. (10 Marks)
b. Explain the measures for selecting best split with example. (06 Marks)

OR

- 8 a. Consider a training set that contains 60 positive examples and 100 negative examples, for each of the following candidate rules.
Rule R_1 : Covers 50 positive examples and 5 negative examples
Rule R_2 : Covers 2 positive examples and no negative examples.
Determine which is the best and worst candidate rule according to
- Rule accuracy
 - Likelihood ratio statistic
 - Laplace measure.
- b. What is Bayes theorem? Explain how it is used for classification.

(09 Marks)

(07 Marks)

Module-5

- 9 a. Explain K-means algorithm. (07 Marks)
b. Write Agglomerative Hierarchical clustering algorithm and explain the single link and complete link techniques. (09 Marks)

OR

- 10 a. Explain DBSCAN algorithm. (08 Marks)
b. Explain CLIQUE algorithm. (08 Marks)
