

CBCS SCHEME

USN

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

18CS641

Sixth Semester B.E. Degree Examination, Feb./Mar. 2022 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 multi-tier architecture with neat diagram. (10 Marks)
b. Give any ten comparisons between OLTP and OLAP. (10 Marks)

OR

- 2 a. Explain with suitable examples :
i) Star schema
ii) Snowflake schema
iii) Fact constellation schema. (10 Marks)
b. Explain the suitable example the various OLAP operations. (10 Marks)

Module-2

- 3 a. Explain with example different methods of indexing the OLAP data. (10 Marks)
b. What is data mining? Explain the challenges that motivate the development of data mining. (10 Marks)

OR

- 4 a. Describe the various types of attributes and data sets. (10 Marks)
b. List and explain the steps involved in data preprocessing. (10 Marks)

Module-3

- 5 a. State Apriori principle. Write Apriori algorithm and apply to following data with support = 50%

Data :

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

- b. Describe alternative methods for generating frequent itemsets. (10 Marks)

OR

- 6 a. Explain the FP – Growth algorithm and construct the FP-Tree for the following :

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}

(10 Marks)

- b. List and explain the measures used in evaluation of association patterns. (10 Marks)

Module-4

- 7 a. Explain decision tree algorithm with example. (10 Marks)
b. Explain the characteristics of decision tree induction algorithm. (10 Marks)

OR

- 8 a. Write K-nearest neighbor algorithm. Explain how it is used in classification technique. (10 Marks)
b. How Baye's theorem can be used for solving a classification problem? Explain with example. (10 Marks)

Module-5

- 9 a. What is Cluster Analysis? Explain desirable features of clustering. (10 Marks)
b. Briefly explain the different methods of clustering. (10 Marks)

OR

- 10 a. Explain the following :
i) K-means clustering (10 Marks)
ii) Graph based clustering (10 Marks)
b. Explain DBSCAN algorithm. List out the issues in cluster validation. (10 Marks)

* * * * *