Printed Pages – 4 Roll No. : .....

C033635(033)

B. Tech. (Sixth Semester) Examination, April-May 2022

(AICTE Branch)

(IT Branch)

DATA MINING

Time Allowed : Three hours

Maximum Marks : 100

Minimum Pass Marks : 35

Note : *Attempt all questions. Part (a) is compulsory from each unit & carry 4 marks. Solve any* ***two*** *from (b), (c) & (d) of each questions and carry 8 marks.*

Unit-I

1. (a) Distinguish between OLTP and OLAP.

   (b) Discuss the tasks of data mining with suitable examples.

(c) What is ETL? Explan the steps in ETL.

(d) What is data transformation and data discretization? Write down the steps involved in data transformation.

Unit-II

2. (a) What is Association Analysis?

(b) Develop the Apriori algorithm for generating frequent itemset.

(c) Explain multidimensional association rule mining in detail.

(d) Write an explain the algorithm for mining frequent itemsets without candidate generation.

Unit-III

3. (a) What are the two techniques of supervised learning?

(b) With an example explain various attribution selection measures in classification.

(c) What are regression problem in supervised learning? Explain with example.

(d) How do you analyze discriminant analysis with suitable steps?

Unit-IV

4. (a) What is Outlier Detection?

(b) Explain the different types of data in cluster analysis.

(c) Explain the partitioning methods and hierarchical methods in brief.

(d) Cluster the following eight points (with (x, y) representing locations) into three clusters :

$A_1(2, 10)$, $A_2(2, 5)$, $A_3(8, 4)$, $A_4(5, 8)$, $A_5(7, 5)$, $A_6(6, 4)$, $A_7(1, 2)$, $A_8(4, 9)$

Initial cluster centers are : $A_1(2, 10)$, $A_4(5, 8)$ and $A_7(1, 2)$.

The distance function between points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as

$$P(a, b) = |x_2 - x_1| + |y_2 - y_1|$$

Use K-means algorithm to find the three cluster centers after the second iteration.

Unit-V

5. (a) What are different types of web mining?

   (b) Explain how data mining is used for retail industry.

   (c) How will you solve a classification problem using Bayesian classification.

   (d) What is the difference between data mining and web mining?