



Reg No.: \_\_\_\_\_

Name: \_\_\_\_\_

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**

B.Tech Degree S8 (R,S) Exam April 2025 (2019 Scheme)

**Course Code: CST466****Course Name: DATA MINING****Max. Marks: 100****Duration: 3 Hours****PART A***Answer all questions, each carries 3 marks.***Marks**

- |    |  |     |
|----|--|-----|
| 1  | List and explain any two applications of data warehouse.   | (3) |
| 2  | Differentiate OLAP and OLTP.   | (3) |
| 3  | Explain the two sampling methods used in data reduction.   | (3) |
| 4  | Calculate the normalized value for 300 in the following set of data points using min-max and z-score normalization: 100, 200, 300, 500, 900. | (3) |
| 5  | Explain splitting indices in a Decision tree. Give examples.   | (3) |
| 6  | Explain the importance of SLIQ algorithm.  | (3) |
| 7  | Describe any three methods to improve the efficiency of Apriori algorithm.   | (3) |
| 8  | Define support, confidence and frequent item set in Association rule mining context.   | (3) |
| 9  | Briefly explain web mining taxonomy.   | (3) |
| 10 | Differentiate between web content mining and web structure mining.   | (3) |

**PART B***Answer any one full question from each module, each carries 14 marks.***Module I**

- |    |  |     |
|----|--|-----|
| 11 | a) Explain the different data mining functions.                        | (7) |
|    | b) Explain three-tier data warehouse architecture with a neat diagram. | (7) |

**OR**

- |    |  |     |
|----|--|-----|
| 12 | a) Explain the differences between star schema and snowflake schema in a data warehouse with a suitable example.                           | (6) |
|    | b) Illustrate different OLAP operations in multidimensional data model with examples. List the differences between ROLAP, MOLAP and HOLAP. | (8) |

**Module II**

- |    |   |     |
|----|---|-----|
| 13 | a) Explain any two data preprocessing steps in detail.  | (8) |
|    | b) Suppose that the data for analysis includes the attribute <i>age</i> . The <i>age</i> values | (6) |



For the data tuples are (in increasing order) 4,8,9,15,21,21,24,25,26,28,29,34.

Use smooth by bin means and bin boundaries to smooth the above data

Where number of bins are 3.

**OR**

- 14 a) Describe the various techniques for numerosity reduction in data mining. (7)  
 b) Why do we need data transformation? What are the different ways of data transformation? (7)

**Module III**

- 15 a) A data base contains 80 records on a particular topic of which 55 are relevant to a Certain investigation. A search was conducted on that topic and 50 records were retrieved. Of the 50 records retrieved, 40 were relevant. Construct the confusion matrix and calculate the precision, recall and specificity scores for the search. (7)  
 b) Explain SLIQ algorithm with a suitable example. (7)

**OR**

- 16 a) Explain partition clustering and the importance of PAM method. (6)  
 b) Explain ROCK algorithm with a suitable example. (8)

**Module IV**

- 17 a) A database has six transactions. Let min\_sup be 33.33% and min\_conf be 60%. (8)

TID	ITEMS
T1	Cake, Bread, Jam
T2	Cake, Bread
T3	Cake, Coke, Chips
T4	Chips, Coke
T5	Chips, Jam
T6	Cake, Coke, Chips

Find frequent item set using Apriori algorithm and generate strong association rules from the dataset.

- b) Explain partition algorithm with a suitable example. (6)

**OR**

- 18 a) Consider a dataset containing transactions from a retail store. The dataset (8)

Consists of the following transactions:

Transaction1:{milk, bread, eggs}

Transaction2:{milk, bread, butter}

Transaction3:{bread, butter}

Transaction4:{milk, bread}



Transaction5:{bread, eggs}

Using the Frequent Pattern Growth Algorithm, find all frequent item sets with

A minimum support of 40%. Then, determine the association rules with a

Minimum confidence of 60%.

- b) Illustrate the working of Pincer Search Algorithm with an example. (6)

**Module V**

- 19 a) Explain web usage mining applications and activities. (7)

- b) Explain context focused crawler and Personalization. (7)

**OR**

- 20 a) Describe different Text retrieval methods. Explain the relationship between text mining, information retrieval and information extraction. (6)

- b) Explain how web structure mining is different from web usage mining and web content mining? Write a CLEVER algorithm for web structure mining. (8)

\*\*\*\*