

**GUJARAT TECHNOLOGICAL UNIVERSITY****BE - SEMESTER– VI (NEW) EXAMINATION – WINTER 2021****Subject Code:3160714****Date:02/12/2021****Subject Name:Data Mining****Time:10:30 AM TO 01:00 PM****Total Marks: 70****Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Simple and non-programmable scientific calculators are allowed.

**MARKS**

- Q.1**
- |     |   |           |
|-----|---|-----------|
| (a) | Justify the importance of data mining.  | <b>03</b> |
| (b) | Differentiate OLTP and data warehouse.  | <b>04</b> |
| (c) | Briefly discussed steps of KDD process. | <b>07</b> |

- Q.2**
- |     |  |           |
|-----|--|-----------|
| (a) | Explain data reduction and dimensionality reduction?   | <b>03</b> |
| (b) | What do you mean by correlation analysis? Justify its importance.  | <b>04</b> |
| (c) | List common task involved in the data pre-processing. Explain briefly any four tasks of data pre-processing with suitable example. | <b>07</b> |

**OR**

- |     |  |           |
|-----|--|-----------|
| (c) | Define the following:<br>concept description, support, confidence, strong association rules, data generalization, and unsupervised learning. | <b>07</b> |
|-----|--|-----------|
- Q.3**
- |     |   |           |
|-----|---|-----------|
| (a) | How the classification is differs from the prediction? Explain phases of classification.  | <b>03</b> |
| (b) | Attribute income have minimum value of 12000 INR and maximum value of 98000 INR. Normalize income value of 73600 INR,<br>(i) Using min-max normalization in the range of [0,1]<br>(ii) Using z-score normalization. Take mean value of income as 54000 and standard deviation is 16000. | <b>04</b> |
| (c) | Using Apriori algorithm, find all frequent itemsets for following transaction data.<br>( Take min_sup=60% and min_conf=80% )  | <b>07</b> |

ID	Items
1	{M,O,N,K,E,Y}
2	{D,O,N,K,E,Y}
3	{M,A,K,E}
4	{M,U,C,K,Y}
5	{C,O,O,K,I,E}

**OR**

- Q.3**
- |     |  |           |
|-----|--|-----------|
| (a) | What is the use of proximity measures? Explain any one proximity measures with equation.   | <b>03</b> |
| (b) | Explain Bayesian learning and inference with suitable example.   | <b>04</b> |
| (c) | List the accuracy parameters used for the performance evaluation of classification and discuss any five parameters with appropriate example. | <b>07</b> |
- Q.4**
- |     |   |           |
|-----|---|-----------|
| (a) | Differentiate supervised and unsupervised learning.   | <b>03</b> |
| (b) | Explain logistic regression with appropriate example. | <b>04</b> |

(c) Explain working of decision tree algorithm with suitable example. **07**

**OR**

**Q.4** (a) Differentiate agglomerative and divisive methods of clustering. **03**

(b) What do you mean by perceptron? Discuss single-layer and multi layer perceptron. **04**

(c) Explain K-means clustering algorithm and prove that outlier adversely affect the performance of algorithm. **07**

**Q.5** (a) Give strength and weakness of k-means in comparison of k-medoids algorithm. **03**

(b) What is outlier? Why outlier mining is important? **04**

(c) Write about different clustering approaches with their strength and weakness. **07**

**OR**

**Q.5** (a) Briefly explain the spatial data mining and temporal mining. **03**

(b) Discuss any four data mining features available in the WEKA. **04**

(c) How data mining is useful for web mining. Discuss any four web mining applications. **07**

\*\*\*\*\*