GUJARAT TECHNOLOGICAL UNIVERSITY
BE- SEMESTER–VI (NEW) EXAMINATION – WINTER 2024

Subject Code:3163209   **Date:09-12-2024**
Subject Name:Data Mining and Business Intelligence
Time:02:30 PM TO 05:00 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)**	In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.	**03**
	(b)	A data warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data – Justify.	**04**
	(c)	Briefly discussed steps of KDD process.	**07**
Q.2	**(a)**	Define: Support, Confidence, Confusion Matrix	**03**
	(b)	Explain Star, Snowflake, and Fact Constellation Schema for Multidimensional database.	**04**
	(c)	With the help of a neat diagram explain the 3-tier architecture of a data warehouse.	**07**

<div align="center">OR</div>

	(c)	List common tasks involved in data pre-processing. Explain briefly any four methods of data pre-processing with suitable example.	**07**
Q.3	**(a)**	How classification differs from the prediction? Explain phases of classification.	**03**
	(b)	What is Decision Tree? Explain how classification is done using decision tree induction.	**04**
	(c)	Consider the following set of transactions. Let min_sup = 30% and min_conf = 60%.	**07**

TID	Items bought
T1	pen, pencil
T2	book, eraser, pencil
T3	book, chalk, eraser, pen
T4	chalk, eraser, pen
T5	book, pen, pencil
T6	book, eraser, pen, pencil
T7	ink, pen
T8	book, pen, pencil
T9	eraser, pen, pencil
T10	book, chalk, pencil

1. Find all frequent itemsets using Apriori algorithm.
2. Generate strong association rules.

<div align="center">OR</div>

Q.3 **(a)** Describe the different classifications of Association Rule Mining. **03**

(b) Apply Min-Max normalization to scale the data into the range [1, 10] to the age values: 10, 30, 45, 23, 57, 63, 72, 27, 37, 55, 15, 32 and the normalized value for age = 30 is _____. **04**

If Mean sales value is Rs. 65,500 and Standard Deviation is Rs. 17,000 then Z-Score Sales value of Rs. 82,200 is _____.

(c) Construct decision tree using ID3 classifier for the given training dataset. **07**

Temperature	Outlook	Humidity	Windy	Played
Mild	Sunny	80	No	Yes
Hot	Sunny	75	Yes	No
Hot	Overcast	77	No	Yes
Cool	Rainy	70	No	Yes
Cool	Overcast	72	Yes	Yes
Mild	Sunny	77	No	No
Cool	Sunny	70	No	Yes
Mild	Rainy	69	No	Yes
Mild	Sunny	65	Yes	Yes
Mild	Overcast	77	Yes	Yes
Hot	Overcast	74	No	Yes
Mild	Rainy	77	Yes	No
Cool	Rainy	73	Yes	Yes
Mild	Rainy	78	No	Yes

Q.4 **(a)** Briefly explain Linear and Non-linear regression. **03**

(b) Explain Bayesian learning and inference with suitable example. **04**

(c) What are the limitations of the Apriori approach for mining? Briefly describe the techniques to improve the efficiency of Apriori algorithm. **07**

OR

Q.4 **(a)** Differentiate OLTP and OLAP. **03**

(b) What is Big Data? Describe 4 Vs of Big Data? **04**

(c) Explain K-Mean clustering algorithm. How K-Mean clustering differs from K-Medoid clustering method? **07**

Q.5 **(a)** Explain NameNode and DataNode in HDFS. **03**

(b) With the help of a suitable example, illustrate the OLAP operations: 'drill-down', 'roll-up', 'slice' and 'dice'. **04**

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

OR

Q.5 **(a)** Explain Click-stream mining. **03**

(b) Briefly explain the spatial data mining and temporal mining. **04**

(c) Draw and explain Hadoop architecture. **07**

****