GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VI (NEW) EXAMINATION – SUMMER 2022

Subject Code:3163209   **Date:14/06/2022**
Subject Name:Data Mining and Business Intelligence
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)	Define: Data, Information and Knowledge.	03
	(b)	Do feature wise comparison between Business Intelligence and Data Warehouse.	04
	(c)	Define KDD. Explain KDD process in detail.	07
Q.2	(a)	What is Market Basket Analysis? Explain Association Rules with two measures: Support and Confidence.	03
	(b)	Differentiate between OLTP and OLAP.	04
	(c)	Discuss research issues in Data Mining.	07
		OR	
	(c)	Explain Mean, Median, Mode, Variance, Standard Deviation and five number summary with suitable database example.	07
Q.3	(a)	Define data cube and explain three operations on it.	03
	(b)	Define noise. Explain binning methods for data smoothing.	04
	(c)	Explain three-tier Data Warehouse Architecture.	07
		OR	
Q.3	(a)	Define: Numerosity Reduction, Data Integration and Data Transformation.	03
	(b)	What is Concept Hierarchy? List and explain types of Concept Hierarchy.	04
	(c)	Define the Apriori Property. Generate candidate itemsets, frequent itemsets and association rules using Apriori algorithm on the following data set with minimum support count is 2 and minimum confidence is 60%.	07

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

			MARKS
Q.4	(a)	Explain Spatial mining using example.	03
	(b)	Define Schema. Explain the following schemas with suitable example. 1) Star 2) Snowflakes 3) Constellations	04
	(c)	Calculate 2 clusters using k-means cluster algorithm. For finding the distance use euclidian distance.	07

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

Assume mean 1 as subject 1 and mean 2 as subject 4.

OR

Q.4	(a)	Explain cluster analysis and outlier analysis with example.	03
	(b)	Discuss the application of data warehousing and data mining.	04
	(c)	Generate decision tree using CART algorithm for the following dataset.	07

Sr. No.	Outlook	Temperature	Humidity	Windy	Class
1	Sunny	Hot	High	False	N
2	Sunny	Hot	High	True	N
3	Overcast	Hot	High	False	Y
4	Rain	Mild	High	False	Y
5	Rain	Cool	Normal	False	Y
6	Rain	Cool	Normal	True	N
7	Overcast	Cool	Normal	True	Y
8	Sunny	Mild	High	False	N
9	Sunny	Cool	Normal	False	Y
10	Rain	Mild	Normal	False	Y
11	Sunny	Mild	Normal	True	Y
12	Overcast	Mild	High	True	Y
13	Overcast	Hot	Normal	False	Y
14	Rain	Mild	High	True	N

Q.5	(a)	Discuss about Big Data.	03
	(b)	Define sampling methods for data reduction.	04
	(c)	What is the need of data preprocessing? Explain data cleaning process for missing values and noisy data treatment.	07

OR

Q.5	(a)	Differentiate Classification and Clustering.	03
	(b)	Minimum salary is Rs. 35,000 and Maximum salary is Rs. 2,10,000. Map the salary Rs. 1,30,000 in new range of Rs. (70,000 , 3,10,000) using min-max normalization method.	04
	(c)	Draw and explain Hadoop architecture.	07

****