Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–V (NEW) EXAMINATION – WINTER 2022

Subject Code:3151608 **Date:04-01-2023**

Subject Name:Data Science

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)**	Why business Analytics is important in now a day?	**03**
	(b)	Differentiate Cross-Sectional, Time Series, and Panel Data.	**04**
	(c)	Explain descriptive, predictive and prescriptive analytics in detail .	**07**
Q.2	**(a)**	Define with example Nominal Scale, Ordinal Scale and Interval Scale.	**03**
	(b)	What are the differences between supervised and unsupervised learning?	**04**
	(c)	Explain significance of Histogram, Skewness and Kurtosis in data analytics.	**07**
		OR	
	(c)	What is Probability Distribution function? Explain Uniform Distribution, Normal Distribution, and Exponential Distribution with suitable scenarios.	**07**
Q.3	**(a)**	Explain Central Limit Theorem	**03**
	(b)	As only 3 students came to attend the class today, find the probability for exactly 4 students to attend the classes tomorrow.	**04**
	(c)	Explain Ensemble Method.	**07**
		OR	
Q.3	**(a)**	Give the difference between Probabilistic Sampling and Non-Probability Sampling.	**03**
	(b)	If a coin is tossed 5 times, find the probability of: (a) Exactly 2 heads (b) At least 4 heads.	**04**
	(c)	Explain Random Forest method.	**07**
Q.4	**(a)**	Explain Chi-Square Automatic Interaction Detection (CHAID) in detail.	**03**
	(b)	How do you calculate maximum likelihood estimation?	**04**
	(c)	Find the least square regression line for the following set of data {(-1 , 0),(0 , 2),(1 , 4),(2 , 5)} b) Plot the given points and the regression line in the same rectangular system of axes.	**07**
		OR	
Q.4	**(a)**	What is Outlier Analysis explain in detail.	**03**
	(b)	Compare linear regression vs. Logistic regression.	**04**

	(c)	Consider the following set of points: {(-2 , -1) , (1 , 1) , (3 , 2)} a) Find the least square regression line for the given data points. b) Plot the given points and the regression line in the same rectangular system of axes.	**07**
Q.5	**(a)**	Explain significance of GINI impurities in splitting dataset.	**03**
	(b)	Explain pros and cons of Decision Tree algorithm.	**04**
	(c)	Explain Decision tree algorithm with suitable example.	**07**

OR

Q.5	**(a)**	Which classification algorithm is preferable when numbers of records are very large, random forest/ decision tree? Justify your answer.	**03**
	(b)	Which are the different matrices to select best model for Classification Problems?	**04**
	(c)	How decision tree and random forest algorithm can be compared on various performances attributes?	**07**

****