Enrollment No./Seat No.:

GUJARAT TECHNOLOGICAL UNIVERSITY
Bachelor of Engineering - SEMESTER - IV EXAMINATION - SUMMER 2025

Subject Code: 3144301 **Date: 12-05-2025**

Subject Name: **Data Science for Humanity**

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.

			Marks
Q.1	**(a)**	Find Q1, Q2, and Q3 for the following data set. {5,40,42,46,48,49,50,50,52,53,55,56,58,75,102}	**03**
	(b)	Draw the framework for data-driven decision making analytics.	**04**
	(c)	Write a short note on web and social media analytics.	**07**
Q.2	**(a)**	Determine the most appropriate data type for the following: a) PIN code b) Exam score out of 100 c) Ranking in a race (1st, 2nd, 3rd…)	**03**
	(b)	Explain measures of shape in brief.	**04**
	(c)	Derive mean of Uniform Distribution.	**07**

OR

	(c)	Derive mean of Poisson Distribution.	**07**
Q.3	**(a)**	In a game, a die is rolled once. What is the probability of rolling a number greater than 4?	**03**
	(b)	A factory has three machines; M1, M2, and M3; producing 30%, 45%, and 25% of the total output, respectively. The defect rates for the machines are: M1: 2% M2: 3% M3: 5% If a product is selected at random and found to be defective, what is the probability that it was produced by machine M2?	**04**
	(c)	Write a short note on Probabilistic Sampling.	**07**

OR

	(a)	A bag contains 5 red balls and 3 green balls. Two balls are selected at random from the bag. What is the probability that the balls selected are red?	**03**
	(b)	In a deck of 52 playing cards, 13 are spades, 13 are hearts, 13 are diamonds, and 13 are clubs. If one card is drawn at random, what is the probability that it is either a heart or a queen?	**04**

	(c)	Explain classical methods of estimation.	**07**
Q.4	**(a)**	Give the difference between classification and regression.	**03**
	(b)	State and explain properties of the Least Squares Estimators.	**04**
	(c)	Explain Naïve Bayes Classifier with suitable example.	**07**

OR

	(a)	Briefly explain application of logistic regression in credit rating.	**03**
	(b)	Explain simple linear regression model.	**04**
	(c)	Explain ID3 algorithm with suitable example.	**07**
Q.5	**(a)**	Define sensitivity. Also, write its formula.	**03**
	(b)	Briefly explain how data science is useful in BioTech sector.	**04**

(c) An educational institution aims to identify students who are at risk of failing a course by analyzing historical data, such as attendance, grades, and participation in extracurricular activities. How would you apply data science techniques to develop a predictive model for identifying at-risk students? Explain the types of data preprocessing techniques you would use, the machine learning algorithms you would consider, and how you would handle class imbalance if more students are predicted to pass than fail. **07**

OR

	(a)	Define Specificity. Also, write its formula.	**03**
	(b)	Briefly explain, how data science is useful for weather forecasting.	**04**

(c) A social media platform wants to perform real-time sentiment analysis on user comments to understand public opinion about a new product launch. The company has access to large volumes of user-generated data and wants to classify comments as positive, negative, or neutral. Describe how you would build a real-time sentiment analysis system using data science. What type of model would you use, and how would you ensure that the system can handle large amounts of data efficiently? Also, discuss the challenges of sentiment analysis in real-time environments and how to overcome them. **07**

***