

GUJARAT TECHNOLOGICAL UNIVERSITY**BE- SEMESTER-IV (NEW) EXAMINATION – WINTER 2024****Subject Code:3144301****Date:21-11-2024****Subject Name:Data Science for Humanity****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) Define data science and data driven decision making.	03
	(b) Explain components of business analytics.	04
	(c) Define descriptive, predictive and prescriptive analytics. Explain predictive analytics in detail with suitable examples.	07
Q.2	(a) Small sample sizes generally provide more accurate estimates of population parameters. True or False? Justify your answer.	03
	(b) Differentiate cross-sectional, time-series and panel data.	04
	(c) Explain importance of skewness and kurtosis in data analytics.	07
	OR	
	(c) Explain importance of measurements of central tendency and data dispersion in data analytics.	07
Q.3	(a) Define discrete and continuous random variables.	03
	(b) Differentiate probabilistic and non-probabilistic sampling.	04
	(c) Discuss Chi-Square distribution, Student's t-distribution and F-distribution.	07
	OR	
Q.3	(a) Define Bayes theorem.	03
	(b) Differentiate method of moment estimation and maximum likelihood estimation.	04
	(c) Discuss binomial, poisson and geometric distributions.	07
Q.4	(a) What is regression? How does it helpful for stock price prediction?	03
	(b) Discuss confusion matrix, accuracy, precision and recall for classification model.	04
	(c) Explain linear regression and estimation of coefficients with a simple example.	07
	OR	
Q.4	(a) How to improve accuracy of regression model?	03
	(b) Discuss optimal cut-off probability and its use.	04
	(c) Explain building decision-tree based classifier with a simple example.	07
Q.5	(a) Tree pruning is a method to avoid underfitting of a decision tree based classifier. True or False? Justify your answer.	03
	(b) Explain random forest in brief.	04
	(c) Discuss how data analytics can be useful for healthcare.	07
	OR	
Q.5	(a) The output of logistic regression is in the range of 0 to 1. True or False? Justify your answer.	03
	(b) Explain Naïve Bayes classifier in brief.	04
	(c) Discuss how data analytics can be useful for education.	07
