

GUJARAT TECHNOLOGICAL UNIVERSITY**BE- SEMESTER-V (NEW) EXAMINATION – WINTER 2024****Subject Code:3151608****Date:09-12-2024****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) What is the purpose of Web and Social Media Analytics? **03**
 (b) Describe Percentile, Decile and Quartile. **04**
 (c) Discuss the framework for Data-Driven Decision Making and Analytics Capability Building. **07**

- Q.2** (a) Differentiate uniform distribution and exponential distribution. **03**
 (b) Differentiate Probabilistic Sampling and Non-Probability Sampling. **04**
 (c) Discuss the importance and application of Probability Density Function (PDF) in continuous random variables. **07**

OR

- (c) The following dataset represents the ages of 10 people: 21, 24, 27, 34, 36, 38, 41, 42, 45, and 50. **07**
- Calculate the mean, median, and mode of the dataset.
 - Calculate the range, variance, and standard deviation.
 - Calculate the skewness of the data and interpret whether it is positively or negatively skewed.

- Q.3** (a) If the probability of picking a red ball from a bag is 0.3 and the probability of picking a blue ball is 0.5, what is the probability of picking either a red or a blue ball? **03**
 (b) Compare linear regression vs. Logistic regression. **04**
 (c) Explain Bayes' Theorem with example. **07**

OR

- Q.3** (a) In a binomial experiment with $n=6$ and $p=0.5$, calculate the probability of getting exactly 4 successes. **03**
 (b) How do you estimate parameters using the Method of Moments? **04**
 (c) Consider the following dataset, which represents the number of hours studied (X) and the corresponding exam scores (Y) of five students: **07**

Hours Studied (X)	Exam Score (Y)
2	50
4	55
6	60
8	65
10	70

Perform basic calculations for linear regression and Predict the exam score for a student who studies for 7 hours.

- Q.4 (a)** Explain the term 'Outlier' in the context of data analysis. **03**
(b) Explain why Central Limit Theorem is called as a heart of the Data Science. **04**
(c) Discuss measure of shapes in detail. **07**
- OR**
- Q.4 (a)** Differentiate regression and correlation. **03**
(b) What is Weight and Bias Tradeoff in Linear Regression? **04**
(c) Explain Maximum Likelihood Estimation (MLE) in detail. **07**
- Q.5 (a)** Explain significance of GINI impurities in splitting dataset. **03**
(b) Discuss the advantages and limitations of Ensemble Methods like Random Forest in classification. **04**
(c) Explain how the Chi-Square Automatic Interaction Detection (CHAID) method works in splitting decision tree nodes. **07**
- OR**
- Q.5 (a)** Which are the different matrices to select best model for Classification Problems? **03**
(b) Explain the categories of Machine Learning Algorithms. **04**
(c) How decision tree and random forest algorithm can be compared on various performances attributes? **07**
