

GUJARAT TECHNOLOGICAL UNIVERSITY**BE - SEMESTER-V (NEW) EXAMINATION – WINTER 2021****Subject Code:3151608****Date:15/12/2021****Subject Name:Data Science****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.

- Q.1 (a)** Define Following Terms: **03**
1. Entropy
 2. Information Gain
 3. Population
- (b)** What are the differences between supervised and unsupervised learning? **04**
- (c)** Compare and Contrast Descriptive Analytics, Diagnostic Analytics, Predictive Analytics, and Prescriptive Analytics with suitable examples. **07**
- Q.2 (a)** Differentiate between univariate, bi-variate, and multivariate analysis. **03**
- (b)** What are dimensionality reduction and its benefits? **04**
- (c)** Explain Following terms with respect to analytics. **07**
1. Mean
 2. Median
 3. Mode
 4. Range
 5. Quartiles
 6. Percentile
 7. Variance
- OR**
- (c)** Explain significance of Histogram, Skewness and Kurtosis in data analytics. **07**
- Q.3 (a)** Explain following terms: **03**
1. Z Score
 2. Normal Distribution
 3. Probability Mass Function
- (b)** What is significance of Poisson Distribution in expectation calculation? Which criteria must satisfy for Poisson Process? **04**
- (c)** What is Probability Distribution function? Explain Uniform Distribution, Normal Distribution, and Exponential Distribution with suitable scenarios. **07**
- OR**
- Q.3 (a)** Define following terms: **03**
1. Standard Error
 2. Sample Mean
 3. Degrees of Freedom
- (b)** Explain Central Limit Theorem. **04**
- (c)** Explain classification of various Sampling methods. **07**

- Q.4 (a)** What is Weight and Bias Tradeoff in Linear Regression ? **03**
(b) Compare and Contrast Linear Regression vs Logistic Regression. **04**
(c) The values of x and their corresponding values of y are shown in the table below **07**

x	0	1	2	3	4
y	2	3	5	4	6

- a) Find the least square regression line $y = ax + b$.
 b) Estimate the value of y when $x = 10$.

OR

- Q.4 (a)** What is significance of Confusion matrix in Model Validation ? **03**
(b) Which are the different matrices to select best model for Classification Problems? **04**
(c) Explain Accuracy, Precision, Recall, F1-Score using following Confusion Matrix **07**

Logistic Regression N=100		Predicted Class	
		False(0)	True(1)
Actual	False(0)	30	20
	True(1)	10	40

- Q.5 (a)** Explain significance of GINI impurities in splitting dataset. **03**
(b) Explain Pros and Cons of Decision Tree Algorithm. **04**
(c) How to Build Decision Tree, given a dataset? **07**

OR

- Q.5 (a)** A cancer detection dataset is used for building classification model and model performs at accuracy of 95 percent. Is this a good model to deploy in real world usage? **03**
(b) Explain various Attribute Selection Measures. **04**
(c) How decision tree and random forest algorithm can be compared on various performance attributes ? **07**

downloaded from
 StudentSuvidha.com