Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–V (NEW) EXAMINATION – SUMMER 2021

Subject Code:3151608 Date:07/09/2021

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.**

Q.1	**(a)**	Explain the measures of central tendency.	**03**
	(b)	Differentiate Stratified Sampling and Cluster Sampling.	**04**
	(c)	Explain the framework for SLR model development.	**07**
Q.2	**(a)**	List various Predictive and Prescriptive analytics techniques.	**03**
	(b)	Explain the measures of shape- Skewness and Kurtosis.	**04**
	(c)	Black boxes used in aircrafts are manufactured by three companies A, B and C. 75% are manufactured by A, 15% by B, and 10% by C. The defect rates of black boxes manufactured by A, B, and C are 4%, 6%, and 8%, respectively. If a black box tested randomly is found to be defective, what is the probability that it is manufactured by company A?	**07**
		OR	
	(c)	Discuss the various measures used to validate the simple linear regression.	**07**
Q.3	**(a)**	What is Non-Probability Sampling? Explain with an example.	**03**
	(b)	List Machine Learning Algorithms and explain any two in brief.	**04**
	(c)	Explain the term: Classification and Regression Tree.	**07**
		OR	
Q.3	**(a)**	Discuss pyramid of analytics with diagram	**03**
	(b)	Define the following terms in Logistic Regression.: I. Sensitivity, II. Specificity	**04**
	(c)	Explain Central Limit Theorem (CLT) in detail.	**07**
Q.4	**(a)**	Explain various needs of business analytics in today's scenario.	**03**
	(b)	Briefly discuss about population and sample.	**04**
	(c)	Explain Random Forest method.	**07**
		OR	
Q.4	**(a)**	List out the criteria used by decision tree to develop the tree.	**03**
	(b)	How the random variables are classified? Discuss in brief.	**04**
	(c)	A hospital is interested in estimating the average time it takes to discharge a patient after the clearance (discharge note) by the doctor. Calculate the required sample size at a confidence of 95% and maximum error in estimation of 5 minutes. Assume that the population standard deviation is 30 minutes.	**07**

Q.5 **(a)** Describe three components of Business Analytics. **03**

(b) Explain the term: Method of Moments. **04**

(c) Time between failures (in hours) of a wire cutter used in a cookie manufacturing oven is given in Table. The function of the wire-cut is to cut the dough into cookies of desired size. **07**

Time between failures of wire-cut (in hours)

2	22	32	39	46	56	76	79	88	93
3	24	33	44	46	66	77	79	89	99
5	24	34	45	47	67	77	86	89	99
9	26	37	45	55	67	78	86	89	99
21	31	39	46	56	75	78	87	90	102

(a) Calculate the mean, median, and mode of time between failures of wire-cuts.

(b) The company would like to know by what time 10% (ten percentile or P_{10}) and 90% (ninety percentile or P_{90}) of the wire-cut will fail?

(c) Calculate the values of P_{25} and P_{75}.

OR

Q.5 **(a)** Discuss the term Bagging. **03**

(b) Describe Gini Impurity Index and Entropy used in CART. **04**

(c) Explain in detail: Maximum Likelihood Estimation (MLE). **07**

****