Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–IV (NEW) EXAMINATION – SUMMER 2022

Subject Code:3144301 Date:27-06-2022

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.**
4. **Simple and non-programmable scientific calculators are allowed.**

			MARKS
Q.1	**(a)**	What are various types of data measurement scales? Which operations can be performed on ration scale?	**03**
	(b)	What is Web Scrapping? How is it used? Give one example of it.	**04**
	(c)	Mention at least three applications of Data Science. Explain a case study in any one of the domains.	**07**
Q.2	**(a)**	Differentiate between Parameter and Statistic. What is the role of inferential statistics?	**03**
	(b)	What is the relationship between PDF and CDF of a continuous random variable?	**04**
	(c)	For what purpose analytics process is used? Explain various types of analytical techniques with examples.	**07**
		OR	
	(c)	Demonstrate the Logistic Regression Model's diagnostics in terms of Classification Table, Sensitivity, and Specificity.	**07**
Q.3	**(a)**	How do you find the standard error of a point estimate?	**03**
	(b)	What does the central limit theorem say? Where is it applicable?	**04**
	(c)	What are various means for measures of position and measures of variability. Explain three of each.	**07**
		OR	
Q.3	**(a)**	How to estimate population variance from multiple samples?	**03**
	(b)	How can the shape of a data distribution be measured? Explain them.	**04**
	(c)	Explain any three of the following: 1. Exponential Distribution 2. Chi-Square Distribution 3. Student's t-Distribution 4. F-Distribution	**07**
Q.4	**(a)**	What is the use of maximum likelihood estimation?	**03**
	(b)	What is pruning? What are prepruning and postpruning techniques?	**04**
	(c)	Explain a case study for classification using a decision tree.	**07**
		OR	
Q.4	**(a)**	How are the correlations and relationships related to each other?	**03**
	(b)	What are the advantages and disadvantages of regression models?	**04**
	(c)	Differentiate the three popular decision tress ID3, CHAID, and CART based on various parameters.	**07**
Q.5	**(a)**	What is sampling distribution and why is it important?	**03**

	(b)	In which scenario, logistic regression is preferred? How it works?	**04**
	(c)	What is the difference between population and sample? What are various sampling methods? Explain them with example.	**07**

OR

Q.5	**(a)**	Tickets numbered 1 to 20 are mixed up and then a ticket is drawn at random. What is the probability that the ticket drawn has a number which is a multiple of 3 or 5?	**03**
	(b)	explain the least squares and the fitted model in concern of a linear regression.	**04**
	(c)	In what ways is NB better than other classification techniques? Compare with decision trees.	**07**

****