GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–IV(NEW) EXAMINATION – WINTER 2022

Subject Code:3144301   **Date:14-12-2022**

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)**	Define business analytics and explain the use of it	**03**
	(b)	Define Following Terms: 1. Entropy 2. Percentile 3. Population 4. Quartiles	**04**
	(c)	Compare and Contrast Descriptive Analytics, Diagnostic Analytics, Predictive Analytics, and Prescriptive Analytics with suitable examples.	**07**
Q.2	**(a)**	Define continuous random variable and discrete random variable with example	**03**
	(b)	Two packets containing chocolates are placed on a table. The packets are labeled P1 and P2. Packet P1 contains 7 strawberry chocolates and 4 orange chocolates. Packet P2 contains 3 strawberry chocolates and 10 mango chocolates. The Packets are arranged so that the probability of selecting packet P1 is 1⁄3 and the probability of selecting Packet P2 is 2⁄3. Aashish is blindfolded and asked to select a chocolate. He will win a color TV if he selects strawberry chocolate. What is the probability that Aashish will win the TV (that is, he will select a strawberry chocolate)?	**04**
	(c)	Explain significance of Histogram, Skewness and Kurtosis in data analytics	**07**
		OR	
	(c)	Explain the framework for SLR model development.	**07**
Q.3	**(a)**	Explain following terms: 1. Cumulative Distribution Function (CDF) 2. Normal Distribution 3. Probability Mass Function	**03**
	(b)	Differentiate Stratified Sampling and Cluster Sampling	**04**
	(c)	Explain Binomial Distribution and Poisson Distribution.	**07**
		OR	
Q.3	**(a)**	Explain why Central Limit Theorem is called as a heart of the Data Science.	**03**
	(b)	What are the differences between supervised and unsupervised learning?	**04**
	(c)	Explain the importance of Sensitivity and specificity in logistic regression Classification Table	**07**

Q.4 **(a)** What is Weight and Bias Tradeoff in Linear Regression ? **03**
   (b) How do you calculate maximum likelihood estimation? **04**
   (c) Explain Random Forest method. **07**

OR

Q.4 **(a)** Explain significance of GINI impurities in splitting dataset **03**
   (b) Which are the different matrices to select best model for Classification Problems? **04**
   (c) The daily footfall at a retail store in Ahmedabad over the last 30 days is shown in Table. **07**

232	277	261	173	283	197	251	212	213	213
229	164	219	196	186	247	244	269	272	216
252	314	161	165	221	260	219	290	225	251

   1. Calculate the mean, median, mode and standard deviation.
   2. Calculate the values of first quartile and third quartile.
   3. Are there any outliers in the data?

Q.5 **(a)** Discuss the term Bagging **03**
   (b) Compare linear regression vs. Logistic regression. **04**
   (c) How to Build Decision Tree, given a dataset? **07**

OR

Q.5 **(a)** What is Outlier Analysis explain in detail **03**
   (b) Explain Chi-Square Automatic Interaction Detection (CHAID) in detail. **04**
   (c) Explain Accuracy, Precision , Recall, F1-Score using any Confusion Matrix **07**

****