

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER-V (NEW) EXAMINATION – WINTER 2023

Subject Code:3155202

Date:22-12-2023

Subject Name:Machine Learning using Python

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) Summarize central limit theorem. State its importance.	03													
	(b) Define the term Machine Learning. Discuss various types of machine learning with suitable example.	04													
	(c) <table border="1"><tr><td>X</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>Y</td><td>11</td><td>12</td><td>13</td><td>15</td><td>17</td><td>19</td></tr></table>	X	1	2	3	4	5	6	Y	11	12	13	15	17	19
X	1	2	3	4	5	6									
Y	11	12	13	15	17	19									
	Find the linear regression coefficient and the model for given data.														
Q.2	(a) Illustrate binomial distribution with example	03													
	(b) Write the limitations of decision tree? How Random forest overcomes that?	04													
	(c) Illustrate in detail: Logistic regression .	07													
	OR														
	(c) Write short note on Gaussian distribution.	07													
Q.3	(a) Differentiate: Supervised vs. unsupervised learning.	03													
	(b) Articulate the challenges in unsupervised learning.	04													
	(c) Discuss any three array creation methods in numpy.	07													
	OR														
Q.3	(a) Explain random variables and their importance.	03													
	(b) Discuss bias-variance trade off in machine learning.	04													
	(c) Write python code to train, test and evaluate SVM classifier using scikit learn library.	07													
Q.4	(a) Differentiate: Discrete distribution vs. continuous distribution.	03													
	(b) Explain K-nearest neighbor algorithm.	04													
	(c) Why preprocessing is important? Discuss various methods of preprocessing.	07													
	OR														
Q.4	(a) Summarize characteristics of K-means algorithm.	03													
	(b) Discuss any one algorithm to handle curse of dimensionality reduction.	04													
	(c) Illustrate the working of DBSCAN algorithm with suitable example.	07													

- Q.5** (a) Explain k-fold cross validation with suitable diagram. **03**
 (b) Discuss the application for each: 1). Need of minimizing false positive, 2). Need of minimizing false negative. **04**
 (c) Consider following confusion matrix: **07**

		Actual	
		+	-
Predicted	+	48	3
	-	2	47

For the given confusion matrix, compute following metric: Recall, Precision, Accuracy, True Negative Rate, False Positive Rate, F1 Score.

OR

- Q.5** (a) Explain TF-IDF. **03**
 (b) Discuss Stemming and lemmatization. **04**
 (c) Explain the text normalization process with example. State its importance. **07**
