

GUJARAT TECHNOLOGICAL UNIVERSITY**BE - SEMESTER-V EXAMINATION – SUMMER 2025****Subject Code:3155202****Date:20-05-2025****Subject Name:Machine Learning using Python****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.

		MARKS
Q.1*	(a) Define machine learning and explain its applications across various domains.	03
	(b) Define random variables and explain their significance in statistical modeling for machine learning.	04
	(c) Describe the primary functionalities of the pandas library and how it facilitates data manipulation and analysis for machine learning applications.	07
Q.2	(a) Explain the role of NumPy in Python for data manipulation and numerical computing in the context of machine learning.	03
	(b) Differentiate between supervised and unsupervised learning.	04
	(c) Provide examples of real-world applications where Python, NumPy, pandas, and scikit-learn are used together for machine learning tasks.	07
	OR	
	(c) Explain the concept of Bayesian inference and its significance in probabilistic reasoning for machine learning tasks.	07
Q.3	(a) Define reinforcement learning and explain its fundamental principles.	03
	(b) Differentiate between discrete and continuous probability distributions, providing examples of each.	04
	(c) What is the k-nearest neighbors (k-NN) algorithm, and how does it work in classification and regression tasks? How do you select the value of 'k' in k-NN?	07
	OR	
Q.3	(a) How might reinforcement learning contribute to advancements in fields such as healthcare or finance?	03
	(b) Describe the difference between linear regression and logistic regression, and provide examples of scenarios where each would be applicable.	04
	(c) Compare and contrast the strengths and weaknesses of linear regression, logistic regression, SVM, decision trees, random forests, and k-NN algorithms.	07

- Q.4** (a) Explain the benefits of using cross-validation compared to a simple train-test split approach for model evaluation. **03**
- (b) What is dimensionality reduction, and why is it useful in unsupervised learning tasks? Discuss the concepts of feature extraction and principal component analysis (PCA). **04**
- (c) Provide examples of real-world applications where linear regression is suitable and where logistic regression is more appropriate. **07**

OR

- Q.4** (a) What is cross-validation, and why is it important in machine learning model evaluation? **03**
- (b) Explain the importance of preprocessing and scaling in unsupervised learning. What are the different kinds of preprocessing techniques, and why are they necessary? **04**
- (c) Explain how ensemble methods like random forests combine multiple decision trees to improve predictive performance and reduce overfitting. **07**

- Q.5** (a) Explain the concept of confusion matrices and how they are used to calculate various evaluation metrics for classification tasks. Give small example. **03**
- (b) Explain how NLP techniques can be applied for sentiment analysis. What are some typical steps involved in sentiment analysis using NLP? **04**
- (c) Explain the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. How does it identify clusters based on density and proximity? **07**

OR

- Q.5** (a) What are the fundamental concepts of Natural Language Processing (NLP), and why is it important in machine learning and artificial intelligence? **03**
- (b) Discuss the role of machine learning algorithms in text classification tasks within NLP. **04**
- (c) Discuss evaluation metrics tailored for multiclass classification problems, including macro- and micro-averaged precision, recall, and F1-score. **07**
