

GUJARAT TECHNOLOGICAL UNIVERSITY

BACHELOR OF ENGINEERING – SEMESTER 5 – EXAMINATION – SUMMER 2026

Subject Code: 3155202

Date: 30/05/2026

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.**

- Q.1**
- (a) Explain k-Fold Cross-Validation technique. **03**
 - (b) Define clustering and its types. **04**
 - (c) Apply PCA to reduce dataset dimensionality and visualize results. **07**
- Q.2**
- (a) Define Machine Learning and list its main types. **03**
 - (b) Describe how bag-of-words model works. **04**
 - (c) Compare supervised, unsupervised, and reinforcement learning with real-life use cases. **07**
- OR**
- Q.2**
- (c) List various Data Cleaning techniques. Explain and write the implementation/demonstration of it using Pandas library. **07**
- Q.3**
- (a) Illustrate key differences between classification and regression. **03**
 - (b) What is the Central Limit Theorem? **04**
 - (c) Implement simple linear regression using scikit-learn. **07**
- OR**
- Q.3**
- (a) Describe features of scikit-learn library. **03**
 - (b) What is NumPy and How is it important? Explain with example. **04**
 - (c) Illustrate the use of probability distributions in spam classification with an example. **07**
- Q.4**
- (a) What is overfitting in supervised learning? **03**
 - (b) Differentiate between linear and logistic regression. **04**
 - (c) Implement a clustering algorithm to group customers with similar purchasing patterns and analyze the characteristics of each segment. **07**
- OR**
- Q.4**
- (a) Discuss the purpose of feature extraction, elaborate with suitable example. **03**
 - (b) Differentiate between DBSCAN and Agglomerative Clustering based on their working principles and applications. **04**
 - (c) Discuss how SVM separates non-linear data using kernel trick. **07**
- Q.5**
- (a) Discuss the challenges in unsupervised learning. **03**
 - (b) Describe metrics for multiclass classification. **04**

- (c) Explain hyperparameter tuning techniques also discuss how to avoid overfitting and underfitting. **07**

OR

- Q.5** (a) Explain ROC curve and AUC in model evaluation. **03**
(b) Explain TF-IDF and its importance. **04**
(c) Implement an NLP pipeline that removes stop-words, performs stemming, and converts a paragraph into its Bag-of-Words representation. Show the output and discuss how it can be used for document similarity analysis. **07**
