

GUJARAT TECHNOLOGICAL UNIVERSITY
BACHELOR OF ENGINEERING – SEMESTER 6 – EXAMINATION – SUMMER 2026

Subject Code: 3164803

Date: 03/06/2026

Subject Name: MACHINE LEARNING

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.**

- Q.1**
- (a)** Define term machine learning with respect to Task(T), Experience(E) and Performance Measure(P) . Explain its types with suitable examples. **03**
 - (b)** Compare Machine Learning with Data Mining and Big Data Analytics. **04**
 - (c)** Describe the major applications of Reinforcement Learning in real-world domains. **07**
- Q.2**
- (a)** Explain the concept of feature selection and discuss subset selection methods. **03**
 - (b)** Describe importance of Feature Redundancy and its advantages in Machine Learning. **04**
 - (c)** A dataset with 5 features is given. Explain step-by-step how PCA reduces dimensions. **07**
- OR**
- Q.2**
- (c)** A dataset with 5 features is given along with class labels for 3 classes. Explain step-by-step how Linear Discriminant Analysis (LDA) reduces dimensions to 2 features. **07**
- Q.3**
- (a)** List and explain different types of regression algorithms. **03**
 - (b)** Explain how logistic regression differs from linear regression. Give one example use case. **04**
 - (c)** Compare Support Vector Machine with Decision Tree based on advantages, limitations, and use-cases. **07**
- OR**
- Q.3**
- (a)** Compare simple linear regression and multiple linear regression. **03**
 - (b)** Discuss the working of the Random Forest classifier with suitable example. **04**
 - (c)** Demonstrate how a decision tree makes classification decisions based on entropy and information gain. **07**
- Q.4**
- (a)** Define clustering. Differentiate between hard and soft clustering. **03**
 - (b)** Explain the working of K-means algorithm with a simple dataset example. **04**
 - (c)** Analyze the challenges faced in anomaly detection using unsupervised learning. **07**
- OR**
- Q.4**
- (a)** Explain the concept of DBSCAN algorithm. How does it identify core, border, and noise points? **03**
 - (b)** Justify the real world applications of hierarchical, density-based (DBSCAN), and partitioning (K-means) clustering algorithms. **04**

(c) Analyze the performance limitations of K-means when applied to non-spherical or overlapping clusters. Propose possible improvements or alternative algorithms **07**

Q.5 (a) List and briefly explain the steps of building a machine learning model pipeline. **03**

(b) Illustrate the concept of bias–variance tradeoff with a clear diagram and practical example. **04**

(c) Design a confusion matrix for a binary classification model of breast cancer dataset. **07**
1. Compute accuracy, precision, recall, and F1-score.
2. Propose how you would tune model parameters or threshold to reduce false negatives without greatly increasing false positives.

OR

Q.5 (a) Describe hyperparameter tuning using grid and random search. **03**

(b) Explain cross-validation and its role in evaluation of performance measures **04**

(c) Construct a confusion matrix for a three-class image classification task (e.g., Cat, Dog, Horse). Based on your matrix: **07**
1. Derive macro-averaged and weighted-averaged F1-scores.
2. Suggest an ensemble strategy to improve these metrics. Justify how your approach will enhance classification performance.
