(T) Roll No. .....

PAPER ID—11444

B. Tech. EXAMINATION, 2025

(Sixth Semester)

COMPUTER SCIENCE AND ENGINEERING

Machine Learning and its Applications

Time : 3 *Hours* *Maximum Marks :* 70

Before answering the question-paper candidates should ensure that they have been supplied to correct and complete question-paper. No complaint, in this regard, will be entertained after the examination.

Note : Attempt *Five* questions in all, Q. No. **1** is compulsory. Attempts *four* more questions by selecting *one* question from each Unit. All questions carry equal marks.

1. (a) Why is data splitting important in Machine Learning ?

(b) Name *two* techniques used to tune hyperparameters of a Machine Learning model.

(c) What is over fitting ? How can you avoid it ?

(d) What are some effective ways to improve the accuracy of a Simple Linear Regression model ?

(e) What is Minkowski distance ?

(f) Explain the importance of finding sensitivity and specificity for machine learning model in healthcare domain.

(g) How does a Support Vector Machine (SVM) separate data for classification ?

$7 \times 2 = 14$

Unit I

2. (a) Differentiate between supervised, unsupervised and reinforcement learning. 7

   (b) Discuss in detail about bias variance tradeoff. 7

3. Describe the Machine Learning life-cycle in detail. Explain each stage along with the challenges faced during implementation of respective stage. 14

Unit II

4. Explain Principal Component Analysis (PCA) and its role in reducing the dimensionality of data. Describe the step-by-step process involved in performing PCA on a dataset. 14

5. Differentiate between feature selection and

feature extraction in data pre-processing. Explain different techniques of feature selection in detailed manner with suitable example. **14**

Unit III

6. A university wants to predict the Final Outcome of its students based on the following attributes : **14**

Study Time (S)	Activity Level (A)	Attendance Quality (Q)	Final Outcome (O)
Short	Active	Poor	Fail
Moderate	Inactive	Poor	Fail
Extensive	Active	Excellent	Pass
Extensive	Inactive	Excellent	Pass
Moderate	Active	Excellent	Pass
Short	Inactive	Poor	Fail
Moderate	Active	Poor	Fail
Extensive	Active	Poor	Pass
Short	Active	Excellent	Fail
Extensive	Inactive	Poor	Pass

Using the ID3 Algorithm, construct a decision tree to predict final outcome (O). Steps to Perform :

(i) Calculate the entropy of the dataset.

(ii) Compute the information gain (IG) for each attribute (Hours Studied, Participation, Class Attendance).

(iii) Choose the attribute with the highest information gain as the root node.

(iv) Recursively split the dataset to build the complete decision tree.

7. (a) Discuss the concept of logistic regression and its use in binary classification. 7

(b) Explain the concept of Naïve Bayes Algorithm with an example. 7

Unit IV

8. (a) Explain the *k*-Means Algorithm with an example. 7

   (b) Compare Boosting and Bagging in ensemble methods, highlighting their key differences in terms of error correction and model performance. 7

9. A company has developed a machine learning model to predict whether a customer will purchase a new product. After testing the model on 100 customers, it was observed that out of 50 actual buyers, the model correctly predicted 40 as 'Yes' and 10 as 'No.' Among 50 customers who were not buyers, the model incorrectly predicted 15 as 'Yes' and correctly identified 35 as 'No.'

Calculate the following performance metrics based on the data provided : **14**

(a) Confusion matrix

(b) Accuracy

(c) Precision

(d) Recall (Sensitivity)

(e) Specificity

(f) F1 Score

(g) Misclassification rate.

