9. Explain briefly the following : 7+7=14

(a) Boosting and Bagging in ensemble methods

(b) Singular Value Decomposition (SVD)

(c) Confusion matrix.

❖❖❖❖❖❖❖

(T) Roll No. .....

PAPER ID—11444

B.Tech. EXAMINATION, 2026

(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, selecting *one* question from each Unit. Q. No. **1** is compulsory. All questions carry equal marks.

(Compulsory Question)

1. (a) What assumption about features is made by the Naïve Bayes classifier ?

   (b) What is the key difference between parametric and non-parametric models ?

   (c) In the *k*-Nearest Neighbours algorithm, what role does the parameter *k* play in classification.

   (d) What is the main objective of K-means clustering in unsupervised learning ?

   (e) What is the significance of the F1-score in evaluating classification models ?

   (f) What is the main purpose of using cross-validation in machine learning experiments ?

   (g) Why are Random Forests generally more robust than a single decision tree ?

7×2=14

Unit I

2. When developing a machine learning solution, several systematic steps are followed to ensure accurate and reliable predictions ? Discuss the machine learning model development process with proper justification for each stage. **14**

3. (a) Explain, how key features of machine learning makes it suitable for solving modern data-driven problems ?

   (b) Discuss four different application areas of machine learning and explain how machine learning techniques help improve decisionmaking in those domains ?

7+7=14

Unit II

4. Explain the problems caused by missing, noisy and inconsistent data in machine learning datasets. What preprocessing techniques can be used to manage these issues effectively ? **14**

5. A sports analytics company collects data from athletes to analyse the relationship between weekly training hours and stamina score. The dataset for four athletes is given below :

Athlete	Training Hours (per week)	Stamina Score
A	9	85
B	6	72
C	11	92
D	7	76
E	8	80

Using Principal Component Analysis (PCA), answer the following :

(a) Determine the eigenvalues and eigen vectors of the covariance matrix.

(b) Identify the principal component that captures the maximum variance.

(c) Project the dataset onto the first principal component and explain how this reduces the dimensionality of the data.

7+3+4=14

Unit III

6. A healthcare system uses a logistic regression model to predict heart disease using patient age. The model has : Intercept = –3 and Age coefficient = 0.04. Class labels : Heart Disease (greater than threshold) and No Heart Disease (less than threshold). **5+4+5= 14**

(a) Based on Logistic regression, calculate the probability of heart disease for a patient aged 50 years.

(b) Determine the predicted class of the patient using the threshold 0.5.

(c) Briefly discuss how changing the threshold influences classification decisions.

7. Answer the following with suitable example and mathematical formulations : **5+4+5=14**

(a) Explain how SVM constructs an optimal separating hyperplane.

(b) Discuss the concept of margin maximization.

(c) Explain the role of kernel functions in handling non-linear classification problems.

Unit IV

8. A regression model is used to predict house prices (in ₹ lakhs). For six houses, the actual prices are : 45, 50, 55, 60, 65, 70 and the predicted prices are : 47, 48, 60, 58, 66, 75.

Answer the following :

(a) Compute the prediction error for each observation.

(b) Calculate the Median Absolute Deviation (MAD).

(c) Compute the Mean Squared Error (MSE) of the model.

(d) Explain how MSE and MAD differ in measuring model errors, particularly in the presence of large prediction errors.

(e) Discuss how analyzing the distribution of prediction errors can help assess the reliability of a regression model.

2+3+3+3+3=14