2025

MCA 3rd Semester Examination

MCA

Paper : MCA - 309

(Machine Learning Lab)

(Practical)

Full Marks : 50 Time : Two Hours

Answer any *one* question on lottery basis : 35×1=35

Write the source code and input / output for each of the programs :

1. Write a program in python jupyter notebook to implement the Naïve Bayes classification algorithm, import all necessary packages and load a data in .csv form either from url or from sklearn datasets. Split the datasets in 75% as training and 25% as test. Find the accuracy of your model.

2. Perform **Principal Component Analysis (PCA)** on the **Wine Dataset** (sklearn.datasets) to reduce its dimensionality. Visualize the data in a 2D scatter plot using the first two principal components and evaluate how dimensionality reduction affects model performance.

3. Use the **Heart Disease Dataset** (available online) to train a **Random Forest Classifier.** Tune the number of estimators and max depth using Grid Search or Random Search, and compare its performance with a single Decision Tree.

4. Build a **Decision Tree Classifier** using the **Titanic Dataset.** Preprocess the data (e.g., handle missing values, encode categorical features) and visualize the decision tree. Evaluate the performance on the test dataset.

5. Train a **k-NN Classifier** on the **Iris Dataset.** Experiment with different values of k and observe the changes in accuracy.

6. Write a program in python jupyter notebook to implement the cross-validation training in any classification algorithm, import all necessary packages and load a data in .csv form either from url or from sklearn datasets. Find the accuracy of your model.

7. Write a Python program to import and export data using Pandas library functions.

8. Load a real-world dataset (e.g., from Kaggle or UCI Machine Learning Repository) for predicting a continuous variable. Implement **Linear Regression**, evaluate the model using R^2 score and Mean Squared Error (MSE), and plot the predicted vs. actual values.

9. Use the **Pima Indians Diabetes** Dataset (available on Kaggle) to train a **Logistic Regression** model. Evaluate the model using metrics like accuracy, precision, recall, and F1- score.

10. Train an **SVM Classifier** on the **Breast Cancer Dataset** (sklearn.datasets). Compare the performance of different kernels (linear, polynomial, RBF).

11. Write a program in python jupyter notebook to implement the knn classification algorithm, import all necessary packages and load a data in .csv form either from url or from sklearn datasets. Split the datasets in 70% as training and 30% as test. Find the accuracy of your model.

12. Write a Python program to demonstrate various Data Visualization Techniques.

13. Write a program in python jupyter notebook to implement the simple linear regression model for the data points x = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9] and y = [10, 30, 35, 50, 75, 80, 85, 90, 105, 125]. Find the coefficients and plot the data values with best fit line.

14. Write a program in python jupyter notebook to implement the k-means clustering algorithm, import all necessary packages and load a data in .csv form either from url or from sklearn datasets. Find the accuracy of your model.

15. Implement **Polynomial Regression** on a synthetic dataset to demonstrate how it handles non-linear relationships. Visualize the fitted curve.

[Viva-voce-10]

[Laboratory Notebook-5]
