National Institute of Electronics and Information Technology, Calicut
Certified AI Professional
Machine Learning
Assignment 11
(Total Marks 40)

(Submit py or ipynb file. All questions carry 4 marks each)

Instructions
- All are ML Model development. Required data exploration (like printing few rows of data, column names, printing target values etc.) is expected before proceeding to ML model development.
- For questions 1 ,3 & 4 use different datasets. Do not use iris dataset/datasets included in sklearn/any of the datasets used in earlier assignments

Decision Tree

1. Select a random dataset suitable for classification, and develop an ML model using Decision Tree Classifier. Print Accuracy Score, Confusion Matrix and Classification Report

Random Forest

2. Implement a Classifier using Random Forest Classifier for the pimaindians dataset. Print Accuracy Score, Confusion Matrix, Precision & Recall

Naïve Bayes

3. Apply Naïve Bayes Classifier for a random dataset. Print all evaluation parameters.

LDA

4. Select a random dataset and apply LDA for classification. Print confusion matrix, accuracy score, precision & recall (hint: use classification_report functions)

PCA

5. For the pima Indians dataset reduce the number of columns to 3 using PCA and develop an ML model using any one ML algorithm

Pickle

6. Develop a classifier for iris dataset. Store the model to a file using Pickle and write code to load it from file to do prediction for the full dataset and print confusion matrix.

Algorithm Comparison

7. Generate the box plot showing the comparison of cross validation accuracies of pima Indians dataset (for the following algorithms)
   a) KNN
   b) Logistic regression
   c) Naive Bayes
   d) DecisionTree
   e) Random Forest

8. Generate the box plot showing the comparison of cross validated RMSE values for different algorithms for the boston dataset. Apply any 5 regressors including Ridge and Lasso

Gridsearch

9. Apply gridsearch for C values in an SVM classifier for Pima-Indians dataset

Pipeline

10. Apply PCA and KNN on iris dataset using pipeline. Print the accuracy score and confusion matrix for the model.