DEPARTMENT OF INFORMATION TECHNOLOGY, NITK SURATHKAL

LAB ASSIGNMENT-2

IT464: FOUNDATIONS OF MACHINE LEARNING

Write a Python program to answer the following

1. **(i) Apply Decision tree, Naive Bayes classifiers and Random Forest to detect the target as yes or no** (for bank data) and **to detect the diabetes as yes or no** (for diabetes data) using the following data sets. (40M)
   https://www.kaggle.com/datasets/krantiswalke/bankfullcsv
   https://www.kaggle.com/datasets/shashankvichare/diabetes-prediction

   (ii) Test the algorithm's performance on the following test datasets.
   Test Datasets:
   (a) bank-test.xls
   (b) diabetes-test.xls

2. **Apply Decision tree** for the California Housing Dataset to predict the house price and show it. (10M)
   https://www.geeksforgeeks.org/dataset-for-linear-regression/
   Note: Exclude "longitude, latitude and ocean proximity" parameters/variables.

   Compute the price for the **"housing2" test data** using the **trained Decision tree**.

3. **Perform SVM and Bayes classifiers** on the following data to predict credit card fraud. (20M)
   https://www.kaggle.com/datasets/nishipatkar/credit-card-details

   (a) Predict credit card fraud for the test data: creditcard-test.xls

4. Perform KNN Classification **to detect the diabetes as yes or no** (for diabetes data) and to **classify the flower type** (for flower data) using the following data sets. (30M)
   https://www.kaggle.com/datasets/shashankvichare/diabetes-prediction
   https://www.kaggle.com/datasets/arshid/iris-flower-dataset

   (a) diabetes-test.xls
   (b) flower-test.xls