HW5

In HW 5, you will have 3 options.

1. Random Forest VS. SVM
   Given the dataset Letter Recognition dataset in
   https://archive.ics.uci.edu/dataset/59/letter+recognition
   Note that the first column of the dataset is the response variable (i.e., y) (You can download yourself, or use the dataset included in the zip file). First choose two letters for this binary classification problem. Then write a Random Forest and a SVM classifier, tuning the parameters and choosing the best parameters for both models. Using the first 80% data for training and the remaining 20% for testing. Explain your choice of parameters and explain why(we need not only the words but also some evidence). Then compare these two models with performance and run time. Briefly discuss the pros and cons for these two models and the results. Please also include the plot with accuracy vs epochs, to see the training process.

2. Tradition classification models
   We are going to compare the following classifiers Naive Bayes, Logistic Regression, and KNN with the same dataset in option1. Using the first 80% data for training and the remaining 20% for testing. Tuning the parameters and choosing the best parameters for all models. Explain your choice of parameters and explain why(we need not only the words but also some evidence). Then, compare these two models with performance and run time. Briefly discuss the pros and cons for these three models and the results.

3. Clusters
   Use the dataset of Bank Customers which is attached to the assignment file. You can find the description from:
   https://www.kaggle.com/datasets/arjunbhasin2013/ccdata/data
   Create K_Means, Hierarchical, and DBScan clusters. Train the models with the dataset. Choose the best cluster numbers for all three models, compare the performance for all three models. Briefly discuss the pros and cons for these three models and the results.

In all three options, you may process the dataset to deal with the correlated variables, outlier data points, null data, etc. You need to explain why you do this kind of processing. For this assignment you need to write a report. In the report, you need to briefly introduce the dataset, explain your data processing, how to tune the parameters for the model, what is the best model, and compare different models. You need to submit both your code and your report. The report needs to be a pdf file. No code file will be graded as 0, and the code should not in the report.