SKLEARN EXERCISES

In this practice we are going to develop classifiers for a classification problem of your choice. First of all, choose a classification dataset from the UCI Repository of Machine Learning Databases:

https://archive.ics.uci.edu/

Then you must load it with the pandas library. You must clean up and convert any data, as necessary. Once the data is ready for use in a pandas DataFrame object, you must complete the following tasks with the sklearn library:

1) Classify the data with a Naïve Bayes classifier:

https://scikit-learn.org/stable/modules/naive_bayes.html

2) Classify the data with a nearest neighbors classifier:

https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

You must find out the best number of nearest neighbors.

3) Classify the data with a decision tree classifier:

https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html

You must find out the best hyperparameters and plot the best decision tree.

4) Classify the data with a Support Vector Machine classifier:

https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

You must find out the best kernel and hyperparameters.

5) Compare the performance of the previous classifiers. You must report several classification performance measures. You should show the Receiver Operating Characteristic plots for the compared models:

https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html

All these tasks must be carried out using 10-fold cross-validation, so that the performance comparisons are carried out on the test sets:

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

You must compose a document that reports all the obtained results, along with a detailed description and explanation of the experiments and the interpretation of the results.

The document (in PDF format) must be uploaded to the associated virtual campus task.