Objective

The objective is to carry out a quantitative study of a classification task with **Decision Trees** and **ensemble techniques (AdaBoost, Bagging, Random Forests)**.

I need a PDF report of **4 pages maximum** (be concise in the explanations). and also code (Jupyter Notebook or Python file) with no size limitation.

Documentation

You may find it useful to check the following documentation:

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

AdaBoost classifier: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html

Bagging classifier: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html

Random forest classifier: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Methodology

You will use Python and Scikit-learn in this work, and a the small face dataset.

With a base code provided, you are asked to play with variations of the base code **to gain a better knowledge** of how the models behave when varying several parameters.

(You may also check the methodology for digit classification e.g. https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html
or https://www.codingame.com/playgrounds/37409/handwritten-digit-recognition-using-scikit-learn)

For each varying parameter, you are asked the following:

- plot train / test accuracy when parameter varies
- display the training time / inference time
- display the confusion matrix for two configurations: the best one and any other bad configuration

Your report will contain 6 parts:

1. Classification with a Decision Tree classifier

The varying parameter is the maximum depth of the tree.

2. Classification with an AdaBoost classifier

Use a decision tree as the base estimator.

The varying parameters are the maximum depth and the number of estimators.

3. Classification with a Bagging Classifier

Use a decision tree as the base estimator.

The varying parameters are the maximum depth and the number of estimators.

4. Classification with a Random Forest

Use a decision tree as the base estimator.

The varying parameters are the maximum depth and the number of trees in the forest (estimators).

5. Extra analysis (mandatory, not optional!)

In this part, you are free to choose an extra relevant topic for analysis. Be imaginative! (e.g. compare Gini and entropy, vary the training set size, compare with another dataset like digit, MNIST, or FMNIST, etc.)

6. Comments and conclusion

Describe what you've learnt in this work, and give a list of observed pros and cons when using decision trees and each of the ensemble classifiers.

LAST question : what is the difference between a bagging classifier and a random forest classifier ?

Good luck.