Machine Learning module – Python Lab – Exam 22/07/2020

Find the best classification scheme for the included dataset.
The solution must be produced as a Python Notebook.
The notebook must include appropriate comments and must operate as follows:

1. The data are split in two files, load the files and merge them according to the first column, which is the index; both the files contain a "class" column, keep both the columns in the merged file calling them "class_x" and "class_y" (you can use the merge function of pandas dataframes) (4pt)
2. Delete all the rows where *class_x* is different from *class_y*, then drop *class_y* and rename *class_x* as *class* (4pt)
3. Reorder the columns in alphabetical order, but placing the class column as the last one; the cleaned dataframe must be named **df**; show its size and head (4pt)
4. Find the best classification scheme using three classification methods
5. For each classification method find the best parameter setting with cross validation on the *training set* (6pt)
6. For each classification method compute the accuracy and the confusion matrix on the *test set* (4pt)
7. Produce a plot of the accuracies given by the methods attempted (3pt)

Quality of the code: (6pt)

- Include appropriate comments with reference to the numbered requirements
- Useless cells, pieces of code and non-required output will be penalized
    - Remove the code you use for testing and inspecting the variables during the development
- Naming style of variables must be uniform and in English
- Bad indentation and messy code will be penalized

Additional directions, the assignments not compliant with the rules below will not be considered

1. The notebook name must be ***emailusername.ipynb*** in lowercase letters
    a. E.G. if your email is mario.rossi45@studio.unibo.it the notebook filename will be mario.rossi45.ipynb
2. The first cell must contain the student first name, last name and email
3. The solution must directly access the data in the same folder of the notebook
4. Upload the notebook only to EoL

Cooperative work will be **heavily sanctioned**
The candidate can freely access the manuals available online.
The candidate can freely access the teaching materials available in the course website, including the available examples of python notebooks.