Exam for Machine Learning Python Lab

Consider the data file linked here horse-colic.csv, and the description file linked here horse-colic-names.txt.
Develop a Jupyter Notebook according to the directions below. Assume that the dataset is in the same folder as the notebook.

The notebook must include appropriate comments and must operate as follows:

1. Load the data into a data frame with the proper coding of nulls, rename the columns to numbers starting from 1, as shown in the description file.

   Remove columns named `3, 25, 26, 27, 28`.

   Explore the data.

   Show the count of nulls for each remaining column.

   ..... **4pt**

2. Use column `23` as target for classification. Drop the rows where the target is null.

   Impute the nulls on the predicting columns using the column means (you can use the `sklearn.impute.SimpleImputer`)

   Separe the dataset into the *predicting* and *target* parts `X, y`

   ..... **4pt**

3. Train, tune and test a classification model of your choice with *cross–validation*

   Plot the resulting confusion matrix on the test set, normalising for true values.

   Show the values *F1-Score* (macro) and *Cohen Kappa Score* on the test set.

   ..... **6pt**

4. Repeat step 3 with a different classification model.

   ..... **6pt**

5. Choose the best model and comment the result.

   ..... **4pt**

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 penalised
- 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 penalised
- Non generalised solution, such as three sequential statements with the same kind of operation instead of a loop, will be penalised

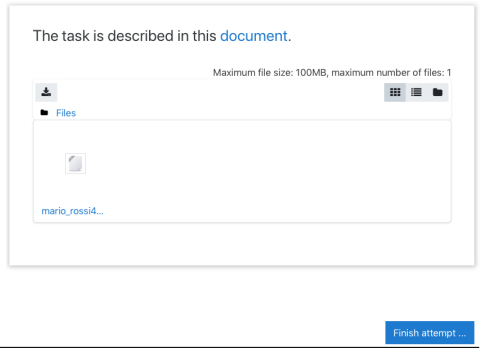
Additional directions, the assignments not compliant with the rules below will not be considered:

- The notebook name must be `youremailusername.ipynb` in lowercase letters (underscore instead of dot)
  E.G. if your email is `mario.rossi45@studio.unibo.it`, the notebook filename will be `mario_rossi45.ipynb`
- The solution must directly access the data in the same folder of the notebook, the name of the file must be the same as the file provided. If the notebook is developed using *Google Colab*, the code must be able to work also out of the Google Colab environment without any change.
- Upload the notebook only to `http://eol.unibo.it` in the activity specified by the teacher, any other way of submitting the notebook will be ignored

Cooperative work will be heavily sanctioned
The candidate can freely access any kind of materials.
See in next page the upload directions.

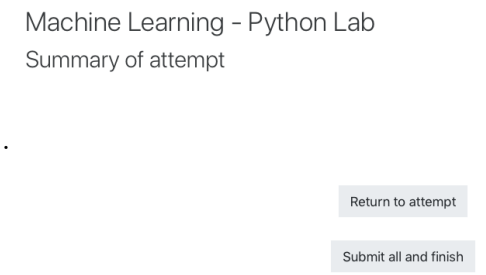
Upload directions Read carefully the requirements, the example and the general directions. Write and test your notebook, save it frequently in your computer. If the program reads a file, do not use a full path, but only the file name.

1. When you have finished, upload your program in the appropriate area .....



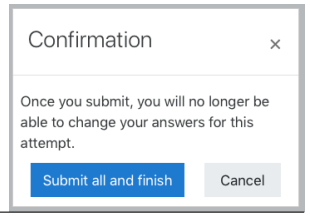
---

2. Push the button Finish Attempt (this action saves your work but still does not submit) .....



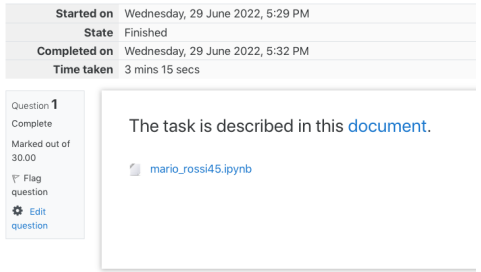
---

3. Push the button Submit all and finish, The confirmation window pops up, ...



---

4. Push again the button Submit all and finish the submission is completed when you see near the top of the window an image like the one at the right.
To observe: **State** Finished
You should see the name of your notebook.
If you see this there is no need to ask for confirmation, if you don't see this, retry the entire process. .....



The submissions which are not properly concluded will not be evaluated

When time expires you are requested to upload your notebook immediately. To withdraw simply finish your attempt according to the procedure above without uploading anything. If you withdraw this attempt will not be considered in any way and will not have any consequence.