Assignment 7

- Please write your code in a Jupyter Notebook file (.ipynb) and submit it on Blackboard
- Make sure to label your answer of each question clearly and add comments to make it readable.
- You are allowed to discuss with other students (up to three) or the instructor. Please put all the names of students that you discussed with. However individual students must write their own solutions.
- Copying a program, or letting someone else copy your program, is a form of academic dishonesty
- Maximally leverage Piazza to benefit other students by your questions and answers.
- Try to be updated by checking notifications in both Piazza and the class webpage.

Use the Bank.csv in attachment. Try the following estimation models and use them to predict customer subscription (y in the table). You should print the cross validation score and confusion matrix for each model. Report which model works best.

- Linear model
- Logit model
- Decision tree
- KNN
- SVM

The description of the dataset is available at:

http://archive.ics.uci.edu/ml/datasets/Bank+Marketing

Notice that you need to pre-process the data. For example, you need to code "no" as 0 and "yes" as 1 in column "loan". Also pay attention "marital" column—it can take three values: single, married and divorced. Can you just code single as 0, married as 1 and divorced as 2?

(Hint: you may find two functions in pandas to be helpful: replace() and get_dummies())

While it is up to you what variables you want to put in the model, you should at least include the following variables:

- Marital
- Education
- Default
- Housing
- Loan
- Duration
- Campaign
- Pdays (the description is incorrect. If value of pdays is -1, it means client was not previously contacted.)