Homework 4

Due 4/25/2020 before midnight

A team project (at most 3 people in a team)

In this assignment, you will perform a machine learning of financial application case study.

You will follow the typical machine learning steps as mentioned in our machine-learning class to analyze a financial problem. Select a suitable Python machine learning module such as sklearn we learned in class or other module(s) that you like to use to perform the machine learning. The following outline the steps that you will performed in this assignment.

1. Determine a financial application that you are interested in.
2. Find and load the dataset.
3. Exploring the data with pandas or other Python modules and do visualizations through plotting so to better know your data
4. Preprocess your data. If needed transform your data (for example, if you use sklearn then converting non-numeric data to numeric data because scikit-learn requires numeric data). If there are missing data fields, decide what to do with those data in this step. For example, remove them or fill them with some default values etc.
5. Decide how to split the data for training and testing
6. Create model
7. Train and test model
8. If you use different machine learning technique to create different models then compare them
9. Tune the model(s) and evaluate performance accuracy. If run different models, then try to explain what and why these model perform differently if you can.

Extra credit: 5 points to 7.5 points

1. If you work alone, you receive 5 extra points automatically
2. If you work as a team, then you will receive (5 * n)/(number of members in the team), where n is the number of different financial applications that your team works on. For example, if your team has 3 members and work on 4 different applications then (4*5)/(3) = 6.7 points
3. The maximum extra credit you can get is 7.5 points. So even if you work alone and work on two different financial applications then you will get min(7.5, (5*2)/1) = min(7.5, 10) = 7.5 points as your extra credit.

Requirements:

1. Use Python
2. Use Jupyter notebook
3. Use Python module(s)

Submission:

1. A MS .doc or .docx file which contains
    a) Student's names (list all the team members' name(s) and RUID(s))

   b) List financial application name(s) and introduction of this application. For example, stock market prediction, housing market prediction, GDP prediction, company profit and loss analysis among many others.
   c) List the link(s) where you get your data and how you download the data. How do you preprocess the data if the data needs to be preprocessed.
   d) List the Python module(s) that you used. And where to download or how to install
   e) List references related to your application(s): website links, journal papers, etc. list reference information separately for different applications.
2. Python notebook(s). Please comment as much as possible inside your notebook. You can use markdown in Jupyter notebook if you know how to use markdown. Or you can comment your steps and code simply just use Python comments. Note one application one notebook. If you work on multiple different applications then submit multiple different notebooks.
3. Create a folder.
   a) Put the report from step 1 into this folder
   b) Put all data files (both downloaded data files before process, and after pre-process if your files need to be preprocessed) that you used in your applications into this folder
   c) Put all notebooks into this folder
4. Zip this folder and submit this zipped file to Blackboard. (if work as a team then each team just needs to submit one copy by any one of the team members)