MFE5310 Assignment 3

Deadline:

- Submit to bb before **23:59:59, May.10, 2020**
- Late submission of an assignment would result in a reduced grade for the assignment, unless an extension has been granted by the instructor. A late submission receives an additional 20 points penalty for every 24 hours delay.

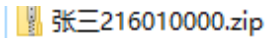
Evaluation

- -20 if your program cannot be executed without any running error
  - write a running guideline if you have multiple files
- all the print messages will be counted to the assignment scores

Submission:

Please hand out a zip file named with **your name and Student No.** which includes your source files of your program executions.

For example,

You should hand out lr.py and compress the them all in a zip file 

.

Refer to the python implementations in our course, write a program in python. The requirements are shown below

1. **Task: predict any financial assets movement direction (up/down) (classification) or return (regression).** No need to be organized as a trading strategy
2. prepare the dataset for any financial assets in China, Hong Kong or US market. **Print** what kind of dataset you are using
   a. Should check if your dataset has class imbalance issue, and **print** the result in your program
3. Implement the machine learning algorithms (hints: try to use keras package for NN model and hmmlearn for HMM)
   a. Convolutional Neural Network (CNN)
   b. Hidden Markov Model (HMM)
   c. Long Short Term Memory (LSTM)
4. **Fine Tune the parameters (hint:** https://machinelearningmastery.com/grid-search-hyperparameters-deep-learning-models-python-keras/**)**
   a. Batch size, number of epochs, optimizer, learning rate, weight initialization for CNN
   b. Batch size, number of epochs, lookback window size (number of neurons, i.e. how many days you need in the past to predict the stock price on the next trading day) for LSTM

   using GridSearch
   a. Normal cross validation
   b. TimeSeriesSplit cross validation
   c. TimeSeriesSplit cross validation with a fixed training size
5. For LSTM, Draw the picture of loss changing with the number of epoch during training. X-axis is the number of Epoch, and Y-axis is the Loss.
6. Find features by yourself. **Print** what kind of features you are using.
   a. Hints: OHLCV
7. Compare the binary classification performance for all the machine learning models you have implemented. Evaluate your results by the metrics below, and **print** it as a pandas data frame
   a. Accuracy

   b. Precision, recall, f1
   c. ROC/AUC
8. **Print** your conclusion
   a. e.g. “LSTM performs the best for my dataset (daily, CSI300 constituents, from 20160801 to 20180901) with the accuracy 75%, precision **, recall **, f1**, auc **”