

## Data process:

1. 將 2009/1/2 ~ 2019/1/2 的資料標上 target,隔天 Close Price 漲則 target=1,隔天 Close Price 跌則 target=0
2. 將資料分為 train(2009/1/2 ~ 2017/12/29)和 test(2018/1/2 ~ 2018/12/31)
3. 新增 year,month,day,dayofweek 4 個 features,並刪除 Date 這個 feature
4. 將 train 和 test 輸出為 csv 檔

## Model:

1. Logistic Regression
2. SVM
3. Neural Network (Dense)
4. Neural Network (LSTM+Dense)

## Predict result:

1. Neural Network (LSTM+Dense) 的準確值是最高的，因為 LSTM 是用在時間序列上的預測，適合用來預測股票
2. 如果 Neural Network (LSTM+Dense)拿來預測其他時間序列的 dataset (ex:天氣)，應該準確率也會是高的

## Improvement:

1. 我在做資料預測前，都有將資料做正規化  
$$\text{train\_x} = \text{train\_x}.\text{apply}(\text{lambda } x: (x - \text{np.mean}(x)) / (\text{np.max}(x) - \text{np.min}(x)))$$
2. 在 Neural Network (LSTM+Dense)裡加入 Dropout 層
3. Neural Network (LSTM+Dense)的 model loss 使用'binary\_crossentropy'，  
binary\_crossentropy 是用來預測 1 or 0，剛好跟我的 target 一樣 (隔天 Close Price 漲則 target=1,隔天 Close Price 跌則 target=0)
4. 因為時間關係沒做，但我覺得可以增加準確度的部分:
  - a. 在 dataset 裡面增加額外的 features (ex:天氣,新聞事件....)
  - b. 調整神經元數量、層數和 learning rate
  - c. Neural Network (LSTM+Dense)的 lookback 我只有用 1 天，可以將 lookback 增加