

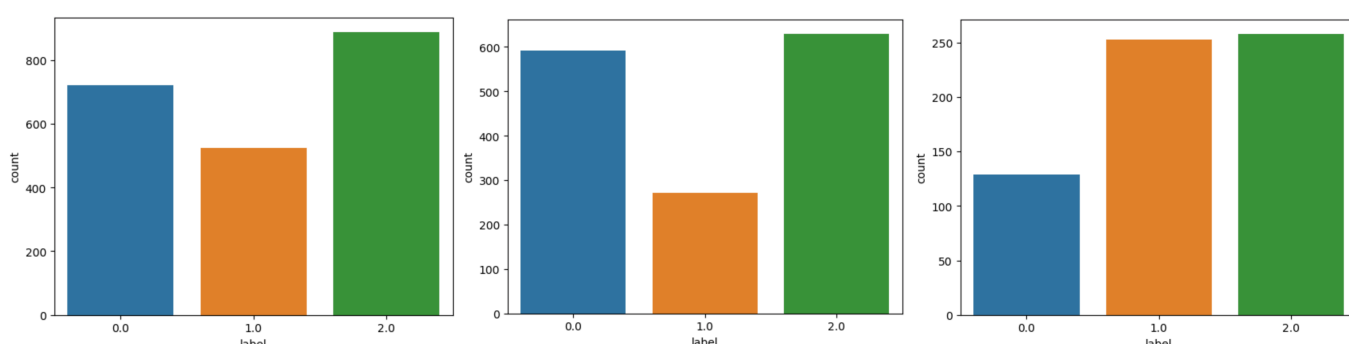
做好的dataset存成dataset.csv放在資料夾裡。

此次作業使用FinMind以及talib完成。

FinMind提供資料集，而talib提供好用的套件。

基本上第一題跟hw2的作法一模一樣，只是改了時間區段而已。

再來是第二題，首先是data distribution，利用seaborn跟pyplot來將每個label的數量印出(從左到右分別是All, Training, Testing):



接著，利用sklearn的RandomForestClassifier以及GridSearchCV，來找出最好的hyperparameter。

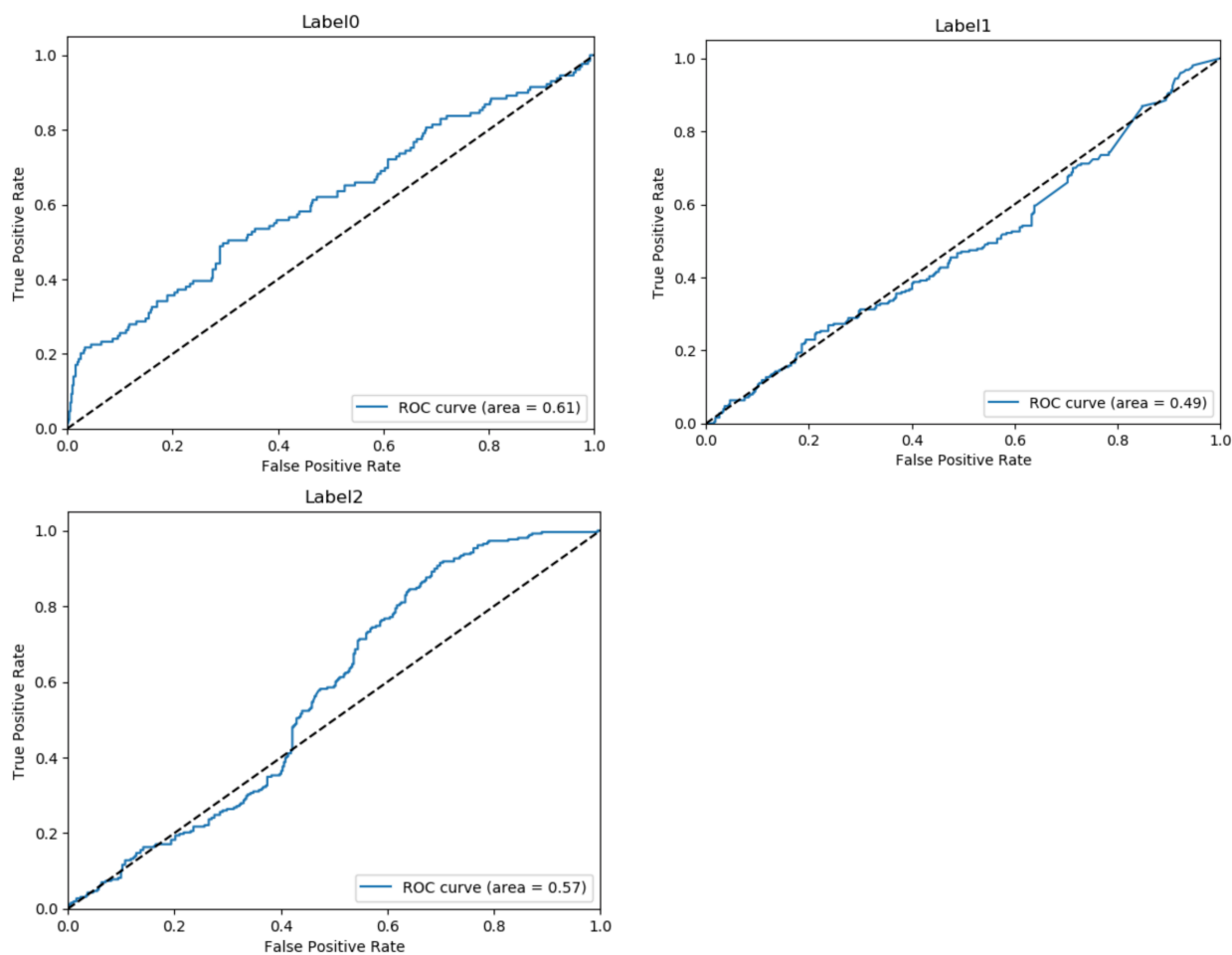
根據助教的setting，找到的best parameter如下圖：

```
{'bootstrap': True,  
 'max_depth': 80,  
 'max_features': 3,  
 'min_samples_leaf': 5,  
 'min_samples_split': 8,  
 'n_estimators': 200}
```

而Training的best score如右： 0.3818210467625027

Testing的準確度如右： 0.4578125

最後是ROC curve，利用sklearn的roc\_curve以及auc完成，將testing的prediction以及testing的ground truth送進function計算，如下圖：



在ROC curve中，在線下的area越大代表預測的越好。