



資料科學與人工智慧競技

Final Project Presentation 2

JPX Tokyo Stock Exchange Prediction

組員：製造所 P96101148 巫清賢
製造所 P96104112 蘇冠瑜

組別證明

Manage Team

Team Name

xian8749

Save Team Name

This name will appear on your team's leaderboard position.

Team Members (2 of 5 maximum)



Su Guan Yu

Member

Make Leader



xian8749 (you)

Leader



JPX Tokyo Stock Exchange Prediction

Explore the Tokyo market with your data science skills

Featured · Code Competition · 1315 Teams · a month to go

611/1315



xian8749

[Add occupation](#)

[Add organization](#)

[Add location](#)

Joined 3 months ago · li

Home

Competitions (1)

Datasets

Code (4)

Account (User ID 9845442)

User Name

xian8749

Your username cannot be changed.

比賽簡介

目標：預測2000支JPX提供股票未來的Change Rate(Target)，以建立投資組合。

Change Rate(Target): 利用明天及後天的Close值計算，公式如右：

$$r_{(k,t)} = \frac{C_{(k,t+2)} - C_{(k,t+1)}}{C_{(k,t+1)}}$$

Target越大表示有利的投資(投資100\$, 得到120\$), 越小則表示不利的投資(投資100\$, 得到80\$)。

評分計算方法為計算Top / Bottom 200的排名分數，公式如右：

$$S_{up} = \frac{\sum_{i=1}^{200} (r_{(up_i,t)} * linearfunction(2, 1)_i)}{Average(linearfunction(2, 1))}$$

簡單而言，就是利用現有資料，找出未來一天的2000支股票的Target值並排序即可完成預測。

資料簡介 - 輸入



stock_prices.csv: 主要訓練資料, 紀錄2000種最常交易的股票, 包含日期、股票編號、開高收低價、交易量、Target等資訊。

secondary_stock_prices.csv: 包含較為冷門的股票數據, 紀錄的資料與stock_price.csv一樣。

option.csv: 紀錄各種股票細節狀態的數據, 包含整天、白天、夜間的開高收低價、結算價、理論價、基本波動率、利率等等。

trades.csv: 上一個營業周的交易量匯總摘要。包含自營交易、證券商交易、個人交易、外匯等的售出額、購買額、總額(購買額+售出額)、差額(購買額-售出額)。

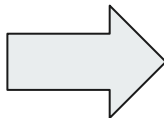
資料分析

1. 將不需要的欄位去除, 以 `features = ['SecuritiesCode', 'Open', 'High', 'Low', 'Close', 'Volume']`, 來預測 `Target`
2. 少部分資料有空值, 需額外進行處理

	Date	SecuritiesCode	Open	High	Low	Close	Volume	Target
0	2017-01-04	1301	2734.0	2755.0	2730.0	2742.0	31400	0.000730
1	2017-01-04	1332	568.0	576.0	563.0	571.0	2798500	0.012324
2	2017-01-04	1333	3150.0	3210.0	3140.0	3210.0	270800	0.006154
3	2017-01-04	1376	1510.0	1550.0	1510.0	1550.0	11300	0.011053
4	2017-01-04	1377	3270.0	3350.0	3270.0	3330.0	150800	0.003026
...
2332526	2021-12-03	9990	514.0	528.0	513.0	528.0	44200	0.034816
2332527	2021-12-03	9991	782.0	794.0	782.0	794.0	35900	0.025478
2332528	2021-12-03	9993	1690.0	1690.0	1645.0	1645.0	7200	-0.004302
2332529	2021-12-03	9994	2388.0	2396.0	2380.0	2389.0	6500	0.009098
2332530	2021-12-03	9997	690.0	711.0	686.0	696.0	381100	0.018414
2332531 rows x 8 columns								

資料分析

```
Date          0
SecuritiesCode 0
Open          7608
High          7608
Low           7608
Close         7608
Volume        0
Target        238
Rank          238
dtype: int64
```



```
Date          0
SecuritiesCode 0
Open          0
High          0
Low           0
Close         0
Volume        0
Target        0
Rank          0
dtype: int64
```

將資料依照SecuritiesCode切成2000個檔案，再對每個檔案中的空值使用平均數(mean)補值。

最後合併2000個檔案依照Date/照SecuritiesCode排序還原成原始資料。

使用方法及介紹



LightGBM

在機器學習中屬於集成學習的方法, 具有以下優點:

- 更快的訓練效率
- 更好的準確率
- 可處理大規模數據
- 支持並行學習
- 低內存使用

輸出結果

- 輸出2021-12-06、07兩天的結果，依 2000支股票代碼排序出它們的Rank

	Date	SecuritiesCode	Rank
0	2021-12-06	1301	733
1	2021-12-06	1332	987
2	2021-12-06	1333	781
3	2021-12-06	1375	434
4	2021-12-06	1376	818
...
3995	2021-12-07	9990	374
3996	2021-12-07	9991	663
3997	2021-12-07	9993	1170
3998	2021-12-07	9994	682
3999	2021-12-07	9997	305

4000 rows × 3 columns

實驗環境



- 作業系統: Window 10
- 程式語言: Python 3
- 使用套件: pandas、numpy、lightgbm、sklearn

結果與討論

1	JPX_LGBM Version 4 (version 4/4) 19 minutes ago by qq21629489 Notebook JPX_LGBM Version 4	Succeeded	0.292	<input type="checkbox"/>
2	JPX_0530 Version 2 (version 2/3) 2 hours ago by Su Guan Yu Notebook JPX_0530 Version 2	Succeeded	0.237	<input type="checkbox"/>
3	notebook9287cd7400 (version 3/5) 2 days ago by xian8749 Notebook notebook9287cd7400 Version 3	Succeeded	-0.024	<input type="checkbox"/>
4	notebook9287cd7400 (version 5/5) 17 hours ago by xian8749 Notebook notebook9287cd7400 Version 5	Succeeded	-0.078	<input type="checkbox"/>
5	notebook02 (version 7/7) 3 hours ago by xian8749 Notebook notebook02 Version 7	Succeeded	-0.142	<input type="checkbox"/>

結果與討論

	1	2	3	4	5
Features	'SecuritiesCode', 'Open', 'High', 'Low', 'Close', 'Volume'				
預測Target	'Target'	'Target'	'Target'	'Rank'	'Target'
備註	有空值時以 mean補值	調整lgbm參數			
預測成績	0.292	0.237	-0.024	-0.078	-0.142



Thanks for listening~