# 見微知著-讓PYTHON成為你的股票理專


- 統計112 林家同
- 資訊112 莊上緣
- 資訊112 李培倫

## Outline

1. Motivations
2. Problem Statement
3. Technical Challenges
4. Dataset to be used
5. Preliminary Methods
6. Evaluation Plans
7. Time Schedule

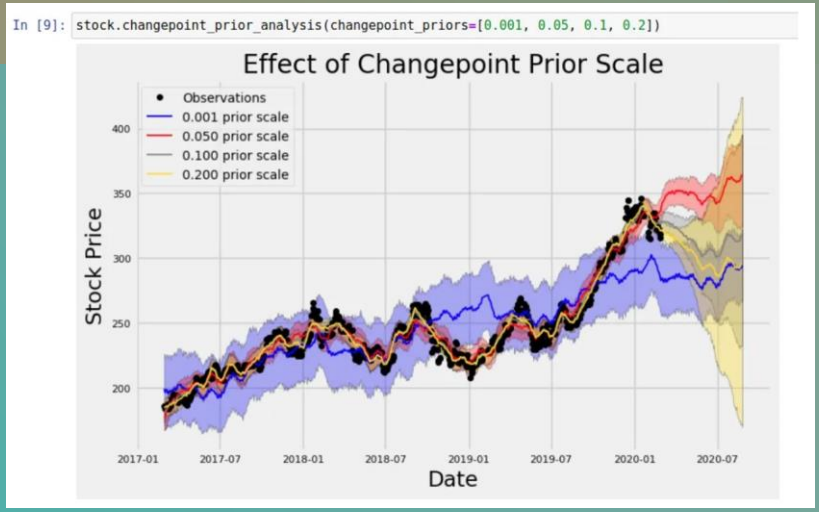
# MOTIVATIONS

1.我們想知道在交易市場，哪些指標會影響股價

2.我們想知道股價的變化波動，是否能找出一些規律

3.調整參數，是否能避免股價預測模型overfitting 或 underfitting

(資料來源: https://weikaiwei.com/finance/stocker/)



# PROBLEM STATEMENT

Input X: 股票代碼
買入股數
賣出時間
Time Interval

Output Y: 推薦度
預估損益

ex: input 2330 10張 2021/01/01-04/30 預計7天後賣出
output 預計報酬率(+58%~-17%)
推薦度(0-100):71
預計損益 +45000

# TECHNICAL CHALLENGES

1. Stock Selection

2. ML Algorithm

3. Feature extraction

# DATASET TO BE USED

- Yahoo Finance
- Taiwan Stock Exchange

```
In [1]: import pandas as pd
        import yfinance as yf
        import matplotlib.pyplot as plt

        stockNo="2330.TW"
        start_date="2021-01-01"
        df=yf.download(stockNo, start=start_date)
        df=df.reset_index()

        [**********100%**********]  1 of 1 completed
```

```
In [2]: df
```

Out[2]:

| | Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|---|
| 0 | 2021-01-04 | 530.0 | 540.0 | 528.0 | 536.0 | 533.814026 | 38770328 |
| 1 | 2021-01-05 | 536.0 | 542.0 | 535.0 | 542.0 | 539.789551 | 34411866 |
| 2 | 2021-01-06 | 555.0 | 555.0 | 541.0 | 549.0 | 546.760986 | 53030554 |
| 3 | 2021-01-07 | 554.0 | 570.0 | 553.0 | 565.0 | 562.695740 | 51166782 |
| 4 | 2021-01-08 | 580.0 | 580.0 | 571.0 | 580.0 | 577.634583 | 59563555 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 83 | 2021-05-17 | 544.0 | 558.0 | 541.0 | 549.0 | 549.000000 | 56270958 |
| 84 | 2021-05-18 | 563.0 | 573.0 | 555.0 | 572.0 | 572.000000 | 43689316 |
| 85 | 2021-05-19 | 571.0 | 572.0 | 565.0 | 567.0 | 567.000000 | 28908777 |
| 86 | 2021-05-20 | 567.0 | 571.0 | 560.0 | 567.0 | 567.000000 | 29709287 |
| 87 | 2021-05-21 | 572.0 | 577.0 | 568.0 | 573.0 | 573.000000 | 27376731 |

# PRELIMINARY METHODS

1.提取特徵值(股價變動斜率、參數...)

2.資料統一化(格式、小數點位數...)

3.測試並找到可能的最佳模型(ML Algorithm、key value...)

# EVALUATION PLANS

loop:

- fetch data
- data normalization
- test
- debug
- if success:
  - return output

# Expected Time Schedule

1.Data Processing (3 days+1~2 days)

2.Algorithm Selecting (1 week+1~2 days)

3.Debugging (3 days+1~2days)

Total: 2~2.5 weeks