STOCK CLOSING PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES[1]

論文復刻及優化

113352011 金碩一　杭佳
113352020 金碩一 涂雅棠
113352033 金碩一 柯明志
113352013 金碩一 蔡博凱
113352001 金碩一 廖皓鈞

11 June, 2025

[1] M. Vijh, D. Chandola and V. D. Tikkiwal, "Stock Closing Price Prediction Using Machine Learning Techniques," Procedia Computer Science, vol. 167, pp. 599–606, 2020.

大綱

01
論文介紹
Introduction

02
研究方法
Methodology

03
實證結果
Empirical Results

04
復刻
Replication

05
延伸討論
Extensions

Data Description & Variables

DataSet

- 資料來源：Yahoo Finance
- 樣本公司：Nike、Goldman Sachs、Johnson & Johnson、Pfizer、JP Morgan
- 期間：2009/04/05–2019/04/05 共 10 年

Table 1. Statistics of the dataset

	Dataset	Training Dataset	Testing Dataset
Time Interval	04/05/2009 – 04/05/2019	04/06/2009- 04/03/2017	04/04/2017 – 04/05/2019

Variables

1. H-L（High – Low）
2. O-C（Close – Open）
3. 7日移動平均（7 Days MA）
4. 14日移動平均（14 Days MA）
5. 21日移動平均（21 Days MA）
6. 7日標準差（7 Days STD DEV）

↓

這些特徵變數會輸入到 ANN 與 RF 模型中，
用來預測次日收盤價。

Random Forest (RF)

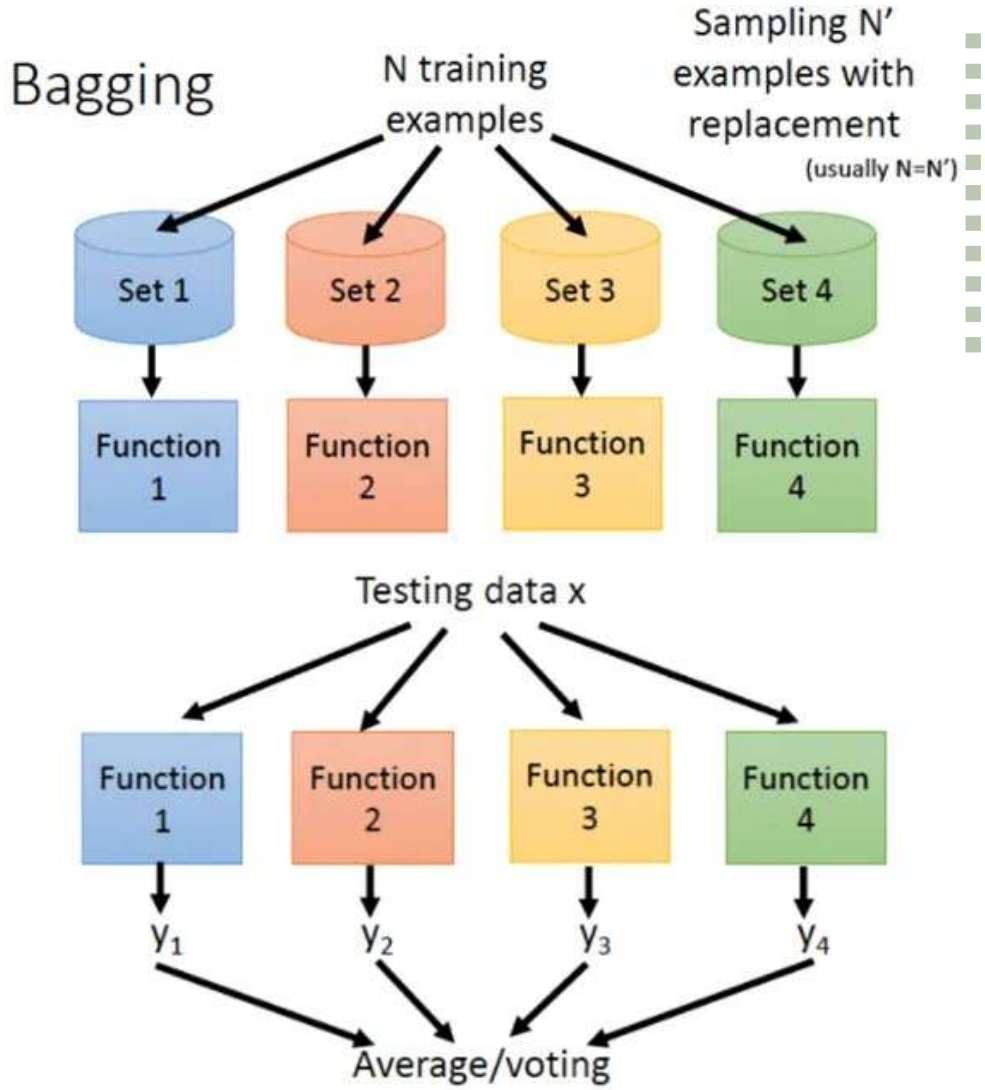
算法演進：

- Bagging(Bootstrap Aggregation)
  - L. Breiman, "Bagging predictors," Machine Learning,vol. 24,no.2,pp.123-140,1996
  - Bootstrap Sampling + Aggregation
- Random Forest
  - L. Breiman, "Random forests," Machine Learning,vol. 45,no.1,pp.5–32,2001
  - Bagging + CART(Classification And Regression Tree)

Random Forest (RF)

Bagging ：

- Bootstrap Sampling
  - 增加隨機性
  - 原始資料集利用拔靴法建立多個子資料集
  - 建立多個學習器(predictor)
- Aggregation
  - 所有學習器的預測值取平均作為最終預測值
  - 多學習器共同決策降低variance
  - 提高泛化能力

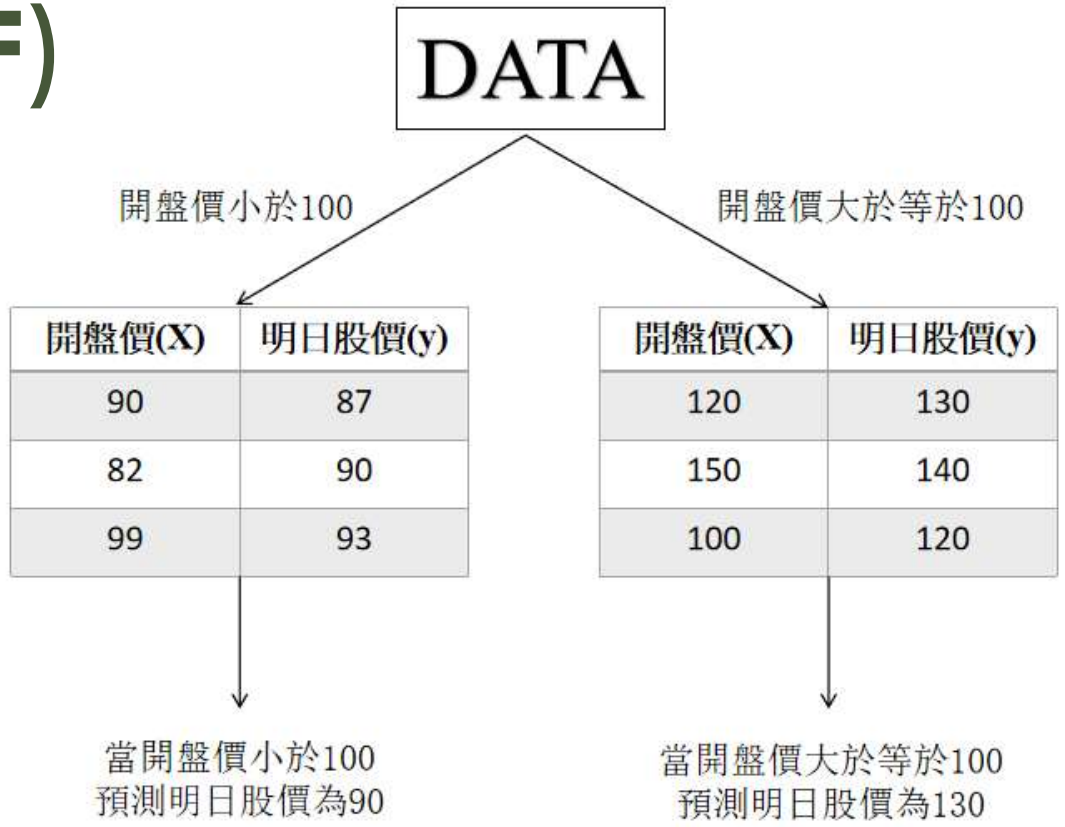


https://medium.com/21-century-girl/random-forest-rf-in-data-analysis-50c3bfa4933a

Random Forest (RF)

Random Forest：

- CART
  - bi-branching by purifying
  - 某特徵可以純化兩邊內部的y
  - e.g.大小讓一籃球成功分成紅球藍球
- Feature selection
  - efficiency + generalization
  - 可自訂分支時考慮的特徵空間
  - 用MSE作為評選特徵的指標

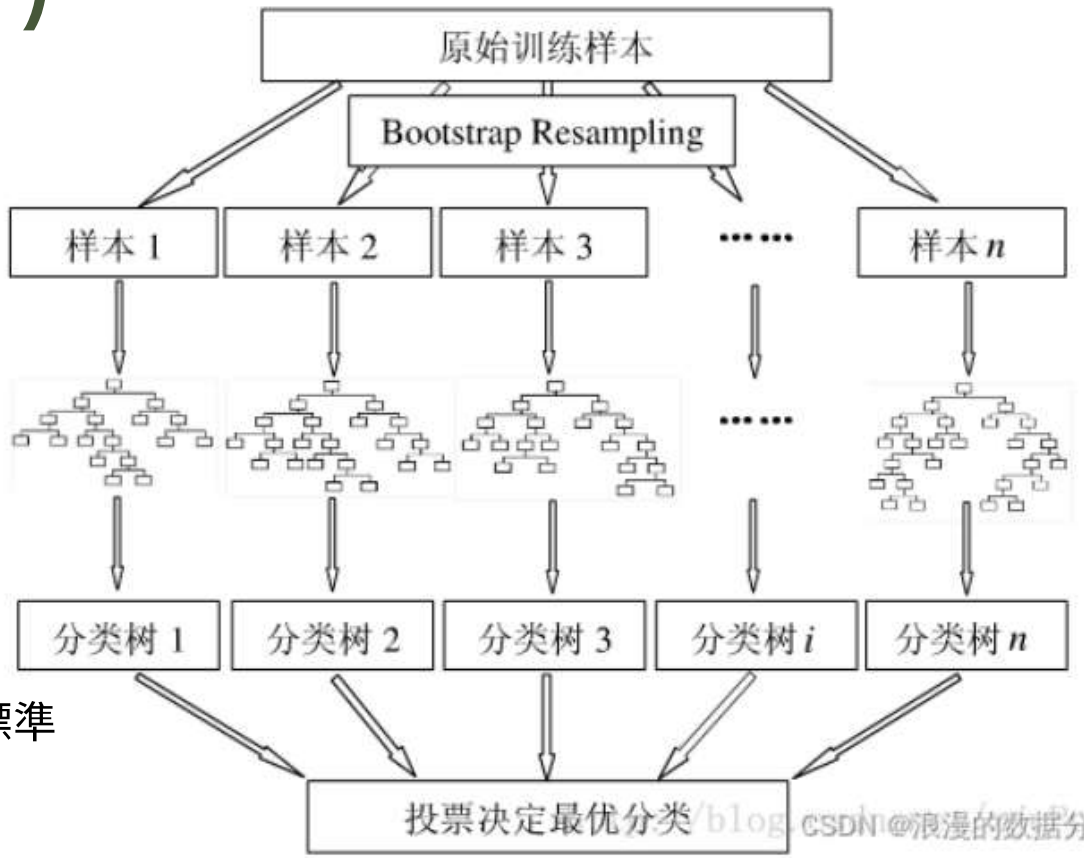


$$MSE(xi) = \sum_{j=1}^{N} (yij - \overline{yi.})^2 , i = 1,2,\ldots,k$$

Random Forest (RF)

超參數設定：

- n_estimators=1000
  - 生成1000顆決策樹
- max_depth=None
  - 不設限樹的深度
- criterion=MSE
  - 結點分裂時使用MSE作為評估特徵的標準
- max_features=auto
  - 結點分裂時的特徵空間包含全部特徵



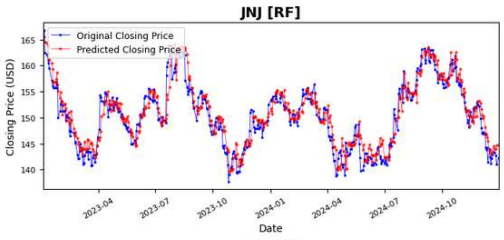
https://blog.csdn.net/weixin_43290383/article/details/123114875

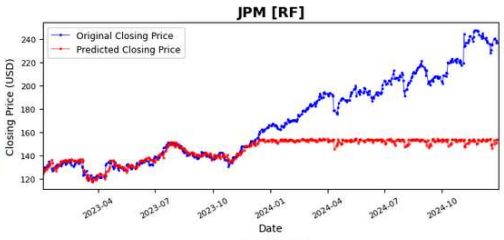
資料來源及區間

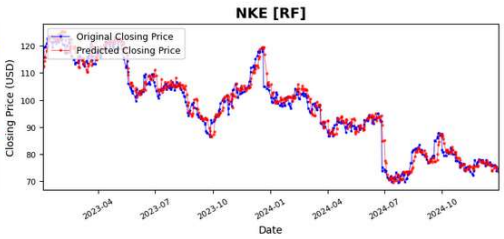
- 資料來源 ： yahoo finance
- 樣本公司 ：JNJ, JP Morgan, Goldman Sachs, Nike, Pfizer (與論文相同)
- 資料區間 ：2015 / 01 / 01 - 2025 / 01 / 01
  - test size ：0.2

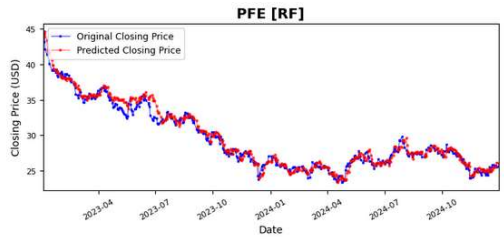
延伸討論 - via return (RF)

via price

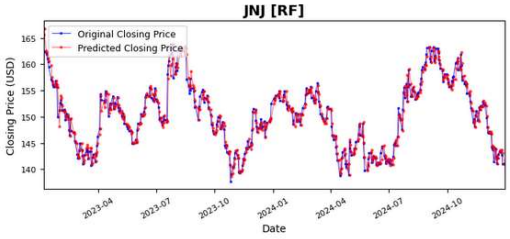








via return



延伸討論

via price	ANN_RMSE	ANN_MAPE	ANN_MBE	RF_RMSE	RF_MAPE	RF_MBE
Compony						
GS	11.9834	2.12%	3.5177	82.5615	10.46%	47.0373
JNJ	2.5474	1.36%	-1.1026	2.5738	1.33%	-0.5132
JPM	4.0624	1.73%	0.5464	38.2917	12.81%	24.8557
NKE	3.6843	2.69%	-0.3684	3.2331	2.44%	-0.4612
PFE	0.7399	1.98%	-0.1055	1.0176	2.47%	-0.3761

via return	ANN_RMSE	ANN_MAPE	ANN_MBE	RF_RMSE	RF_MAPE	RF_MBE
Compony						
GS	6.6117	0.01%	0.7741	8.1449	0.01%	1.5331
JNJ	1.5101	0.01%	0.0037	1.5522	0.01%	-0.0238
JPM	2.5424	0.01%	0.0137	3.3631	0.01%	0.4529
NKE	2.1141	0.01%	-0.2917	1.9142	0.01%	-0.2763
PFE	0.4589	0.01%	-0.0316	0.4447	0.01%	-0.0389

Reference

- [1]M. Vijh, D. Chandola and V. D. Tikkiwal, "Stock Closing Price Prediction Using Machine Learning Techniques," Procedia Computer Science, vol. 167, pp. 599–606, 2020.
- [2] https://online.stat.psu.edu/stat857/node/155/
- [3] https://www.ibm.com/think/topics/ridge-regression
- [4]https://www.youtube.com/watch?v=s9Um2O7N7YM&list=PLXVfgk9fNX2IQOYPmqjqWsNUFl2kpk1U2&index=35
- [5]L. Breiman, "Bagging predictors," Machine Learning,vol. 24,no.2,pp.123-140,1996
- [6]L. Breiman, "Random forests," Machine Learning,vol. 45,no.1,pp.5–32,2001