

★ Project共筆 Project Collaborative Writing ☆

/*The notes for the final project:*/

Proposal 相關

Topic: 股票漲跌預測 (stock price prediction)

Progress

目標goal: 預測台灣0050+0050選的個股共51之股票的買進賣出

Predict buying or sale of the Taiwan 50 and those companies related to Taiwan 50. Total 51.

短期目標 short goal: 分析各指數(基本資料)對於漲跌的影響

Analysis the how the factor relates to the fallen and rise.

Step1: 預測一日漲跌

9/21 HW: understand the meaning of stock indices. (put the notes in 知識相關區)

分工:

抓data (data collection): 友信、宜謙

找paper (paper survey): 彥安、嘉琚、PHU

找source code (code survey): 宛臻、慈娟

程式相關 (about programming)

宛臻

python 網路爬蟲程式的教學

<https://pala.tw/python-web-crawler>

宜謙

台灣股票市場統計資料庫

<http://www.tedc.org.tw/tedc/bank/stock/ch2.htm>

友信

之前我有做過的部分與紀錄, 抓取TWSE的0050股票資訊

[Python_利用Jupyter Notebook抓取 '股票個股日成交資訊' 輸出成csv檔 一步一步圖文操作](#)

慈娟

tensorflow example code - 莫凡

RNN: <https://morvanzhou.github.io/tutorials/machine-learning/tensorflow/5-08-RNN2/>

LSTM: <https://morvanzhou.github.io/tutorials/machine-learning/tensorflow/5-09-RNN3/>

彥安

<https://blog.louie.lu/2017/07/24/%E4%BD%BF%E7%94%A8-python-%E5%A5%97%E4%BB%B6-twstock-%E6%8A%93%E5%8F%96%E5%8F%B0%E8%82%A1%E8%82%A1%E7%A5%A8%E8%B3%87%E6%96%99-01-%E5%A6%82%E4%BD%95%E5%85%A5%E9%96%80/>

10/9 宛臻: github resource

implements and environment of Deep Reinforcement Learning & Supervised Learning for Quantitative Trading (★ 642)

<https://github.com/Ceruleanacg/Personae>

Predict stock market prices using RNN model with multilayer LSTM cells + optional multi-stock embeddings (★ 542)

<https://github.com/lilianweng/stock-rnn>

Use NLP to predict stock price movement associated with news (★ 236)

<https://github.com/WayneDW/Sentiment-Analysis-in-Event-Driven-Stock-Price-Movement-Prediction>

10/11 慈娟: github resource

Plain Stock Price Prediction via RNNs with Graves LSTM unit. (★ 58)

<https://github.com/IsaacChanghau/StockPrediction>

A Stock Analysis and Prediction Toolkit using Additive Models (★ 857)

<https://towardsdatascience.com/stock-prediction-in-python-b66555171a2>

<https://github.com/WillKoehrsen/Data-Analysis/tree/master/stocker>

Stock Market Predictor using Supervised Learning (★ 278)

<https://github.com/scorpionhiccup/StockPricePrediction>

=====

知識相關 (stock knowledge)

9/27

宜謙:

<https://www.youtube.com/watch?v=i5ZJMx5ZS4s>

短線:

1. 十幾萬投下去一次約賺4.5千, 盯著電腦看著股價波動
2. 不要買自己不認識不了解的股票, 要將掌握權握在自己手中
3. 美股帳戶內若沒25000美元, 5天只能做3次交易
4. 短線須做功課, 否則跟賭博沒兩樣 (控制風險)

長線:

1. 巴菲特看好一家公司, 會等股價下跌後再進行購買
2. 長期來看, 總得來說還是會反彈成長 (好公司)
3. 只聽別人建議不做功課, 有90%的人會虧錢

<https://wiki.mbalib.com/zh-tw/%E8%82%A1%E7%A5%A8%E4%BB%B7%E6%A0%BC%E6%8C%87%E6%95%B0>

股票指數:

1. 股票價格指數即股票指數。是由證券交易所或金融服務機構編製的表明股票行市變動的一種供參考的指示數字。
2. 一些金融服務機構就利用自己的業務知識和熟悉市場的優勢, 編製出股票價格指數, 公開發佈, 作為市場價格變動的指標。
3. 股票指數: 表明股票行市變動情況的價格平均數。
4. 以某個基期當作100, 以後的股價和基期做比較, 計算升降百分比。

5.一個較長的時期來說，股票指數比股價平均數能更為精確地衡量股價的變動。

<https://www.cmoney.tw/notes/note-detail.aspx?nid=6494>

投資分析的三大主流:(1) 基本面 (2) 技術面 (3) 籌碼面

1. 基本面：

- 1) 一間公司的股價終究會回歸公司的基本價值。
- 2) 價格始終都圍繞在股票的基本價值上下。
- 3) 價值型投資者會透過公司的財務報表、產業的前景等資訊，評斷出這間公司的內在價值。
- 4) 在股價低於內在價值時買進，直到股價高於內在價值時，再進行賣出。

2. 技術面：

- 1) 研究過去，預測未來。(符合deep learning之分析，是可做的方向)
- 2) 分析市場的供給需求，判斷接下來的市場走向。
- 3) 所有的數據都是以『成交量』和『股價』為基礎。
- 4) 技術指標：均線、RSI、KD
- 5) 一個好的技術分析者，最重要是了解每個指標背後的意義！

3. 籌碼面：

- 1) 找到影響股價的巨人。
- 2) 在市場中，大資金可以影響股價的變動。
- 3) 當主力看好一檔股票，
 - (1) 就是他有信心短期操作這檔股票的波動，
 - (2) 要嘛就是他透過強大的分析，確信這支股票前景看好！

4. 統整：

	基本面	技術面	籌碼面
複雜程度	高	低	中
適用週期	長期投資	短期投資	中短期投資

理財之路: <http://www.cmoney.tw/notes/note-detail.aspx?cid=22814&nid=6637>

(感覺有超多東西，以後慢慢看><)

宛臻：

16種常見股票技術指標：<http://mypaper.pchome.com.tw/tony168324/post/1321691958>

我自己觀察會看以下(但我是買基金)教我賺錢QQ俄可是我也只賺一點點QQ

1. 移動均線MA:短期均線上揚超過長期均線=>進場
否則出場
2. 隨機指標KD:若MA為黃金交叉，且KD高於80必進場
否則會再觀察XDD

通常股市投資分析有三種類型:(1) 技術型 (2) 基本面 (3) 籌碼面

<https://www.cmoney.tw/notes/note-detail.aspx?nid=6494>

通常技術型會挑幾個主要指標觀察，看到買進訊號就買，不過為什麼不一次全部列入考慮呢？
這感覺可以deep learning model來分析，客製化的指標選股(?)

彥安：

1. 股票知識入門:

https://www.youtube.com/watch?v=0ChNmimF6_k

1) 選股:按照風險進場

P.S. 但是我覺得這次專題實在是有點難做出可以自己選股的model

(1) 保險:買大型績優股ex. 鴻海、台積電(但買不起啊問號問號 叫爸爸給你錢), 盈餘每股往往3元以上, 緩緩上升緩緩下滑

(2) 積極:

- 話題股:近期外資、或法人近期積極搶進, 成交量衝高, 股價迅速攀升, 媒體上大量曝光的股票ex.玉晶光

- 概念股:目前具有市場性被民眾認同的一批股票ex.蘋果概念股, 適合短線

(3) 穩健:操作活潑的個股ex.產業龍頭股 統一

2) 基本面:獲利跟本, 籌碼面、技術面:判斷進退場時機

3) 判斷公司

財務狀況:每股盈餘、週轉率、償債能力

營運概況:營運模式、獲利來源、上下供應商、目標市場

競爭同業

替代品

政府政策

4) 多久看盤一次:看股票特性再決定多久看一次盤

2. 自己有稍微看過籌碼面資料, 基本上籌碼面必須要走長期, 不太適合作為此次研究主題 data, 因為要看法人是不是有偷偷跟進, 為了悄悄把股票收回來, 不會一次買太多, 也會製造跌的機會。基本面也有點不適合喔, 因為要看懂財報(還有要看他們有沒有做假帳, 公司營運狀況如何)

做一個抓假帳的model?搶我妹的工作機會(壞 <--壞

慈娟:

1.短線投資技巧:

技術分析是最適合短線投資者使用的方法, 是透過股價的規律推論出的經驗法則認為市場存在規律, 而且歷史會一再重演, 透過技術分析, 可以讓短期的分析有個依據, 提高短線賺價差的勝率

2.技術分析: K線、移動平均線、KD、RSI、MACD、乖離率

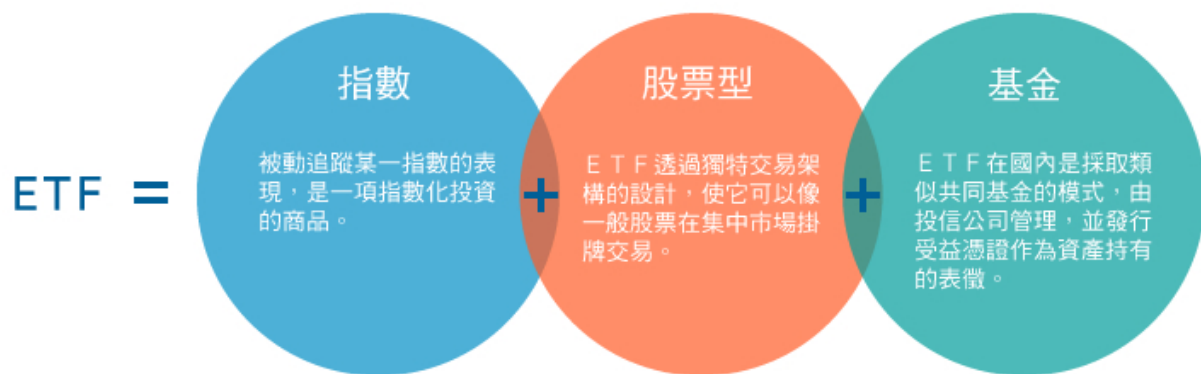
3.目標output: 股市漲跌無變化(股價)(多久預測一次 預測多久後的未來?)

4.可能input: 交易量及金額、歷史股價

5.時間單位? 每小時 每日

友信:

- 何謂ETF?



ETF是英文Exchange-Traded Fund的縮寫，意思是可以在交易所交易的基金，我們如果看交易所提出正式的名稱，應更能理解ETF的真正意涵和這個商品的特色。

ETF交易所正式的名稱是「**指數股票型基金**」，將這個名稱拆成三部分正好可以說明ETF是什麼。

首先是「**指數**」，ETF被設計成被動追蹤某一指數的表現，是一項指數化投資的商品；

其次是「**股票型**」，ETF透過獨特交易架構的設計，使它可以像一般股票在集中市場掛牌交易；

最後是「**基金**」，ETF在國內是採取類似共同基金的模式，由投信公司管理，並發行受益憑證做為資產持有的表徵。

總而言之，**ETF**是被動追蹤某一指數表現的共同基金，其投資組合儘可能的完全比照指數的成分股組成，並且在集中市場掛牌，如同一般股票交易買賣。

ETF示意圖：以台灣科技ETF為例



- 0050 哪50家公司？各公司比例多少？

元大台灣50成分股

股票名稱	持股比例%
台泥	1.21
亞泥	0.82
統一	2.16
台塑	4.6
南亞	4.09
台化	3.11
遠東新	1.14
台肥	0.48
中鋼	2.73
正新	1.42
裕隆	0.58
和泰車	0.76
光寶科	0.79
聯電	1.47
台達電	2.15
2311	1.43
鴻海	7.92

股票名稱	持股比例%
仁寶	1.09
2325	0.92
台積電	18.78
聯強	0.9
宏碁	0.68
鴻準	1.07
華碩	1.9
廣達	1.87
友達	0.67
中華電	3.11
聯發科	2.92
可成	0.94
宏達電	2.16
彰銀	0.58
華南金	0.94
富邦金	1.92
國泰金	2.09
開發金	0.97

股票名稱	持股比例%
元大金	1.26
兆豐金	1.87
永豐金	0.83
中信金	1.84
第一金	0.96
統一超	1.17
大立光	0.57
台灣大	1.6
緯創	0.61
群創	0.54
TPK-KY	0.73
3697	1.01
遠傳	1.09
合庫金	0.79
台塑化	1.51

=====

data相關

TWSE website

<http://www.twse.com.tw/zh/>



美國股票

<https://histock.tw/usstock>



9/28 老師建議:加入新聞等資訊

=====

murmur區

☺ ☹

問題:

一、資料問題

1. 為什麼要拿期貨選擇權的資料
2. 為什麼不抓ETF內的個股資料來分析ETF

二、結果問題

1. 橫軸縱軸代表的意思
2. 結果是表現哪一支ETF?
3. SVM的節果中, y 跟 y_{real} 有什麼不一樣

三、方法問題

1. 為什麼要選那些方法(SVR, XG-boost), 不只用LSTM
2. 如果時間夠的話, 會再加什麼影響因子
3. 為什麼現在好像很少人用deep learning成功預測股市

四、那你們真的有在用那個model投資嗎?

★ 上課筆記 Course Collaborative Writing ☆

/*The notes on the class:*/

9/21上課筆記

- ppt download:

[https://kid.ee.ncku.edu.tw/~course/107-1-big_data/handouts/\[BigData\]200320Introduction20to20Data20Mining20and20Machine20Learning.pdf](https://kid.ee.ncku.edu.tw/~course/107-1-big_data/handouts/[BigData]200320Introduction20to20Data20Mining20and20Machine20Learning.pdf)

彥安/宛臻

1. Association Rules mining

minimum support: 最小出現次數要大於這個才算candidate

minimum confidence $P(D|A)/P(A)$

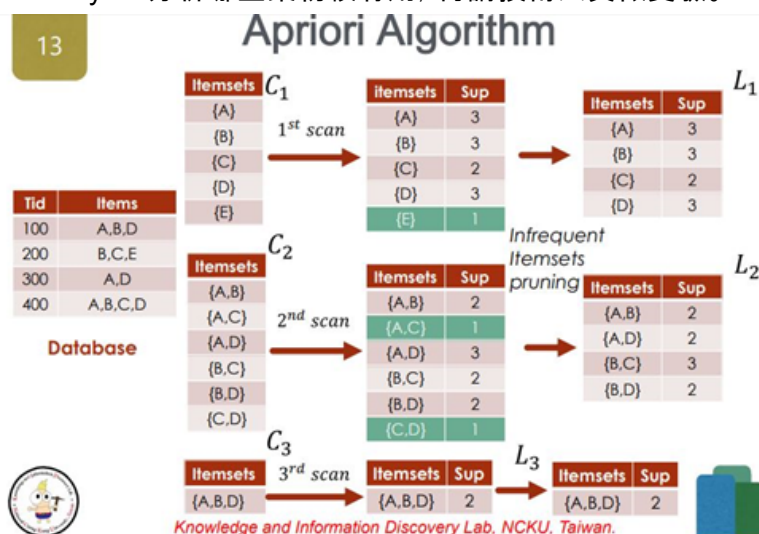
imply: $A \rightarrow D$: $\min_conf = |AD|/|A|$ A為前提下AD出現次數

以上為Apriori algo.

※If the subset doesn't meet the condition, the superset doesn't need to check.

宜謙:

- Big data: 現今出現以前無法處理的資料量 (需要real-time處理)。
- Semistructure: 一半是有structure的資料, 另一半沒有
- Data cleaning: 如何知道資料是突然變化很大, 或是sensor出了問題。
- Data selection: 如何選擇對自己有用的資料。
- Pattern evaluation: 從過去的資料中推測, 重複出現的資料 (ex.溫度超過某值, 濕度超過某值, 那隔天就會下雨。)
- Post processing: 如何呈現找到的資料。
- Data streams: sensor不斷產生的資料。
- Sequence data: 像是DNA資料。
- Graphs: 每個點和邊都有關係。
- Marketing: 人的喜好和產品的相關性。
- Web analysis: 影評 文字 分析推斷電影好看否。
- Recommender system: 推薦前10名電影。
- Bio and med analysis: 分析哪些藥物較有用, 再請技術人員做實驗。



- 主要概念:subset條件不符合, 那就不用檢查superset的條件是否符合
- ◆ 在C1中E的minimum support就不符合, 後面就不用檢查有E的superset。
 - ◆ C2、C3之概念也相同。

9/28上課筆記

彥安

做到18寸的晶圓? 為什麼不做到20寸?

越做越大會因為烘烤有形變, 做久了會形成碗公