

商業分析系統

專題編號：105 -CSIE -S024

執行期限：104 年第 1 學期至 105 年第 1 學期

指導教授：王正豪

專題參與人員： 102590034 顏大竣

一、摘要

建置一個網站，透過連接 python 之後台程式，使用 python 之 requests, BeautifulSoup 工具擷取台灣證券之三大法人之各項資訊，再以程式碼將擷取之資料，連接到 SQLite 建立資料庫，再透過專門分析資料的 R 語言，建立各項圖表，以便預測證券之未來趨勢。

此外，讓建置網頁的可以選擇股市及三大法人證券，在股市分析的方面，參考商科用書，再透過 R 語言建立各項技術指標，再透過隨機森林演算法，確立各項技術指標權重值，再以重要指標建立模型，最後分析該項股票之未來趨勢，以及該次分析的精確度。

關鍵詞：python, R, quantmod, requests。

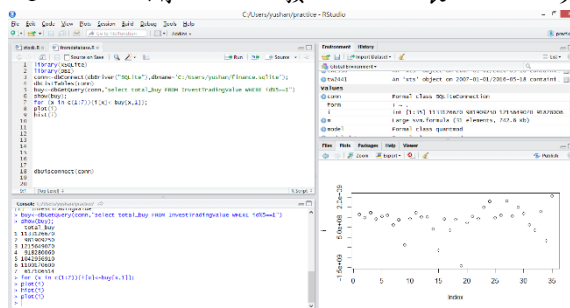
二、緣由與目的

現在世代電腦越來越發達，軟工人員之重要性也日益提高，隨著雲端技術發展，大數據分析也跟著崛起，透過 python 及 R 語言，資料分析也越來越簡單。但是通過分析出來的圖表，也許對於工程師來說，他也不曉得其所代表的意義，但是如果找來商科的專家，卻無法如此有效率的得到他所需要的圖表數據，此實務專題之目的就在於商業人員可以較輕易的得到其所想要的圖表，而理工科的人，可以透過商業的各項技術指標，來決定自己投資的方向

三、研究報告內容

此項專題計畫主要內容為數據分析，透過一些簡單的資料蒐集，發現現在的數

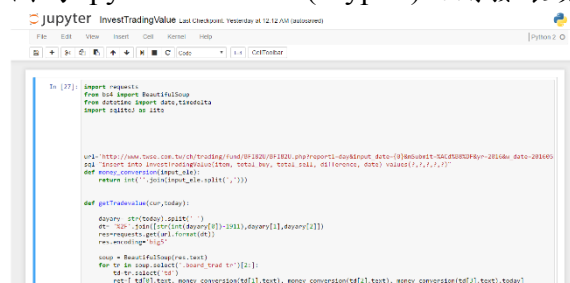
據分析程式有兩種方式，R 及 python，其中 R 比較擅長的地方是分析資料，提供很多的 package 讓開發人員使用，透過 package 裡的函式，輕鬆分解資料、整合資料、繪製圖表，甚至有一些數值方法的計算也可以使用到 R 進行執行，圖為 RStudio 之開發環境



(圖一)RStudio 開發環境

相較之下，python 所擅長的方面就是抓取網頁資料，python 提供之 package 中，很大的部分是讓開發人員可以輕鬆抓取網頁資訊，以及連接各式資料庫建置的系統，通常任何網站，只要沒有防護的建設，透過 python 抓取，程式碼都非常簡單，反而在整理抓取下來的格式會花比較多的時間。

圖為 ipython notebook(Jupyter)之開發環境

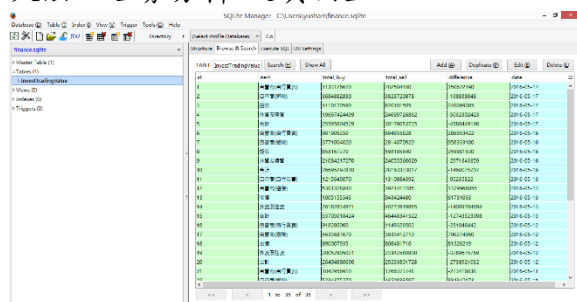


(圖二)ipython notebook 開發環境

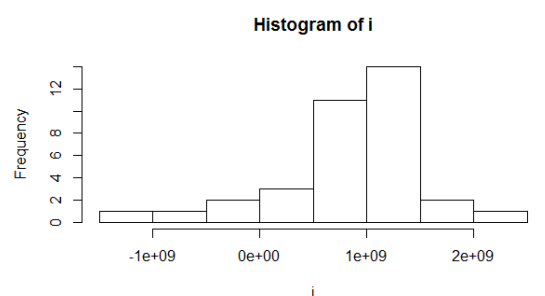
我製作之專題成果主要就是由上述兩種開發環境共同開發，未來我還會加上其他的軟體程式，將這兩項軟體開發的成果進行整合，期許能透過網頁輸入，以查詢各股市之相關資訊，及動態

的計算出其所輸入之公司的股市技術指標，以及給一班投資者建議的投資方向。

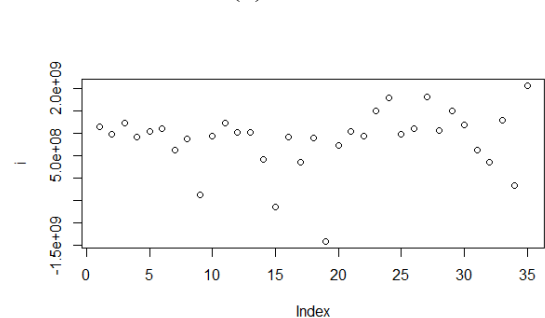
傳送命令至 python 透過 requests 及 BeautifulSoup 擷取三大法人證券之買賣資訊，因此每次進入網站時，都會刷新資訊，再將擷取之資料存入 SQLite(圖三)建置的 database 中，在使用 R 使用 RSQLite package 讀入 database 的資料，將資料分類好，在計算個數學算式(EX:標準差、方差)，最後將其繪製成各項圖表(長條圖、折線圖)(圖四)之後輸出在”三大法人證券分析”之頁面上。



(圖三)SQLite Manager 開發環境



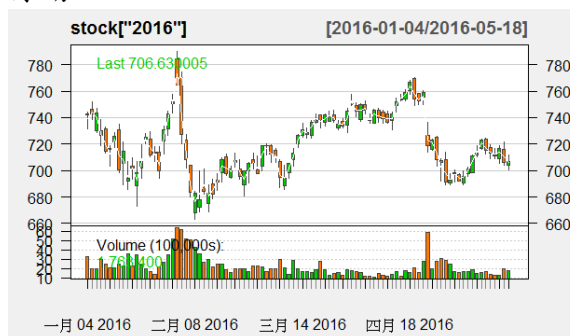
(a)長條圖



(圖四) (b)點圖

在關於股市分析操作，使用者將上市公司代碼輸入後，網頁會將程式傳送到 R 的程式裡，程式會透過 R 裡面的 quantmod package 中的 getsymbols() 抓取股票資料，因為每次抓取都可以抓到最

新的資訊，所以也不用擔心得到的訊息是過時的，在抓取完資料後，會畫出 K 線圖(圖五)，以供投資者參考股市之價格浮動。



(圖五)股市 K 線圖

在繪製完圖形後，將各項資料套入

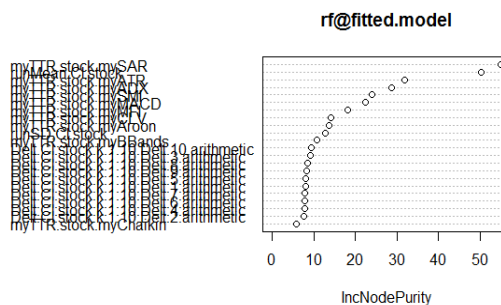
$$P_i = \frac{C_i + H_i + L_i}{3}$$

$$V_i = \left\{ \frac{P_{i+j} - C_i}{C_i} \right\}_{j=1} \quad i = 1, 2, \dots, k$$

其中 P 表示平均價格，V 表示相較前一天之收益百分點，C、H、L 為收盤價、最高價、最低價

在計算 $T = \sum_v \{v \in V_i : |v| > p\%$

在取得 T 值後，計算各項技術指標 (EX:ATR、SMI、ADX、...etc.)，再以 randomForest 方法去掉多餘的變量，後再取各參數值重要程度(圖六)，再將臨界值(本專題取>15 的值)以上之值取出進行函數模型建構，最後將 T 值和決策值歸類”buy”、”hold”、”sell”套入函式模型，就能取得如下圖表(圖七)



(圖六)技術指標之重要程度

```

signal.true
signal.p Buy Hold sell
Buy 11 2 1
Hold 7 39 8
sell 0 5 11
    
```

(圖七)決策結果

最後再將此項決策結果的精確值透

過程式，後再將上面所產稱的圖表及決策結果傳輸至”股市分析頁面”上，讓投資者進行參考

四、參考文獻

- [1] 大數學堂-最好的大數據分析課程網站 <http://www.largitdata.com/>
- [2]韓偉、毛俊杰 著。R 語言與商業智能。
出版地點:電子工業出版社
- [3]醫學統計與 R
<http://web.ntpu.edu.tw/~cflin/Teach/R/R06EN06Graphics.pdf>
- [4]用 R 處理 PostgreSQL/PostGIS 資料(1)
<http://mutolisp.logdown.com/posts/206944-treatment-with-r-postgresql-postgis-data>
- [5]使用 R 的金融統計-數據鋪子
<https://site.douban.com/182577/widget/notes/10568316/note/264808127/>
- [6]jg 經營之家
<http://bbs.pinggu.org/thread-414836-1-1.html>
- [7]python 簡介-廖雪峰의官方網站
<http://www.liaoxuefeng.com/wiki/001374738125095c955c1e6d8bb493182103fac9270762a000/0013747381369301852037f35874be2b85aa318aad57bda00>