# 崑山科技大學 資訊工程系 111 學年度專題製作精簡報告

智慧財報網路平台開發

# SMART FINANCIAL REPORT NETWORK PLATFORM DEVELOPMENT

學 生:資工四A-蔡德龍

資工四A-顏渙儒

資工四B-陳佳賦

指導老師:李宗儒

111年 12月 20日

# 智慧財報網路平台開發之研究

#### 摘要

隨著 AI 人工智慧與機器學習領域的 快速發展,利用程式自動分析上市公司財 報資料與開發演算法挑選優良的公司,成 為在智慧股票分析的熱門主題之一。

本研究以 Python 網路爬蟲技術自動 擷取與分析上市公司財報資料,利用 Data Science 資料科學的分析技術,採用財報 資料近五年 EPS、近五年 ROE、近五年毛利 率、現金殖利率,搭配機器學習方法-支持向量機(Support Vector Machine)與 決策樹(Decision Tree)進行模型訓練,建 構一個以 Python Flask 為 Server 框架的 網頁平台,具有財報數據、線圖,搭配機 器學習模型,提供使用者快速找到推薦的 財報優良公司。

#### **青、前言**

本系統機器學習模型正負樣本訓練資料,是根據上市公司財報資料中的九項數據,四項數據分別為近五年 EPS、近五年ROE、近五年毛利率、現金殖利率,依照符合的指標數量給予 1(正樣本符合數量>=3)或-1(負樣本符合數量<2)。

本系統提供使用者基本的資料建檔,可提供使用者查詢財報資料的搜尋,根據機器學習方式,將四項數據利用支持向量機(Support Vector Machine)與決策樹(Decision Tree)進行模型訓練與測試,利用機器學習模型分析出上市公司中財報優良的公司顯示給使用者,也會將哪些數據不符合以致未納入財報優良公司的原因顯示。

#### 貳、相關研究

在主頁面中,我們透過 JQuery 的技術 JQuery UI AutoComplete 讓使用者搜尋股 票時,自動從資料庫搜尋是否有符合的股 票代碼(圖1.智慧財報分析網路平台開發系統),若有資料,則顯示在下拉式選單中並讓使用者能夠自行選擇股票,並透透的trap Table 讓使用者能夠自由對達大學,並給予單季、近四季、代碼一個股)並給予指標上的說明與公式,在計算指標的過程中,不斷參考財報狗、玩股網、等相關財報網站,進行資料上的設差,讓使用者能對,進而縮小指標上的誤差,讓使用者能對。

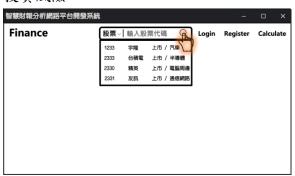


圖 1. 智慧財報分析網路平台開發系統



圖 2. 股票代碼 - 個股

使用者登入自己帳號密碼後,能夠使 用資料庫內的財報計算數值,並可透過產 業別上的按鈕來進行各項指標的排行與歷 史資料的查看(圖 3.股票代碼 - 產業 別)。



圖 3. 股票代碼 - 產業別

本研究以 Python 網路爬蟲技術自動 擷取與分析上市公司財報資料,利用 Data Science 資料科學的分析技術,採用財報 資料近五年 EPS、近五年 ROE、近五年毛利 率、現金殖利率,搭配機器學習方法-支持 向量機(Support Vector Machine)與決策 樹(Decision Tree)進行模型訓練,建構一 個以 Python Flask 為 Server 框架的網頁 平台,提供使用者股票推薦系統。

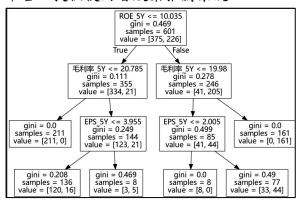


圖 4. 決策樹 Decision Tree

本研究在支持向量機利用 LinearSVC 分類器(如圖.5 LinearSVC 示意圖)解決訓練結果與過度擬合的問題,將參數 C(懲罰係數)調至合理範圍內的數值(參數 C:越大代表在訓練樣本時會愈精準,太大容易過度擬合,太小容易低度擬合),得到的準確度為 98%的訓練結果。

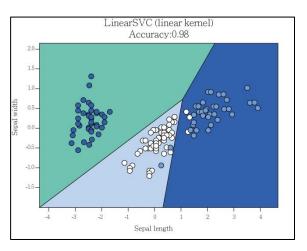


圖 5. LinearSVC 示意圖

本研究利用 Gini index 分類器,由於使用上較為彈性,所以容易有過度擬合 (Overfitting)的問題,預防過度擬合採用了設限的功能,限制樹的最高層數,得到的準確度為 94%的訓練結果。

本研究在機器學習模型的正負樣本訓練資料,是依據公司財報資料中的四項數據分別為近五年 EPS、近五年 ROE、近五年毛利率、現金殖利率,依照符合的指標數量給予1(正樣本符合數量>=3)或-1(負樣本符合數量<2)(如圖 6.正負樣本四項指標)。



圖 6. 正負樣本四項指標

本系統目前可以達到的訓練正確率為 98%(如圖.7訓練結果),測試正確率為 96%, 網頁平台的部分,也提供不同視覺化的呈 現,提供更清晰的分析結果。

```
svc=SVC(C:1000.0)
svc.fit(X.train, y_train)
y_redsvvc.predict(X_test)
print('Robel accuracy_score with rbf kernel and C-1000.0 : {0:0.4f}'. format(accuracy_score(y_test, y_pred)))
linear_svc100-SVK(kernel-'linear', C:100.0)
linear_svc100-SVK(kernel-'linear', C:100.0)
linear_svc100-SVK(kernel-'linear', C:100.0)
linear_svc100-fit(X_train, y_train)
y_pred:linear_svc100-predict(X_test)
print('Model accuracy_score with linear kernel and C:100.0 : (0:0.4f)'. format(accuracy_score(y_test, y_pred)))
print('Training set score: {:.4f}'. format(linear_svc100.score(X_train, y_train)))
print('Training set score: {:.4f}'. format(linear_svc100.score(X_train, y_train)))
Print('Training set score: {:.4f}'. format(linear_svc100.score(X_train, y_train)))
Model accuracy_score with rbf kernel and C-1000.0 : 0.9785
Model accuracy_score with rbf kernel and C-1000.0 : 0.9830
Training set Score: 0.9862)
```

圖 7. 訓練結果

利用系統四項指標設置的標準,將各股票指標進行 0 與 1 的分類(圖 8.模型訓練集),為了避免過度擬合的問題,利用 sklearn模型提供的 train\_test\_split 功能,這個功能能夠隨機分割訓練與測試資料,將資料分割為 70%訓練、30%測試(圖 9.分割、訓練、測試),為了得到更精準的結果,將資料不停訓練與測試,最後得出財報優良公司的結果(圖 10. 模型測試結果)。

	近 <b>5</b> 年平 均本益 比	近5年平 均EPS	近5年 ROE	近5年營 業利益 率	近 <b>5</b> 年營 收成長 率	負債 比率	流動 比率	速動 比率	利息保 障倍數
1101	0	1	0	1	1	1	0	1	1
2330	0	1	1	1	1	0	0	1	1
2344	1	0	0	1	1	1	0	1	1
4207	1	0	0	0	1	1	0	1	1
9943	1	1	1	1	0	0	0	1	1
2303	0	0	0	0	1	0	1	1	1
2409	0	0	0	0	1	1	0	1	0
3231	1	1	0	0	1	0	0	0	1
5276	0	0	0	0	0	0	0	1	1
8112	0	1	1	0	1	0	0	0	1

圖 8. 模型訓練集

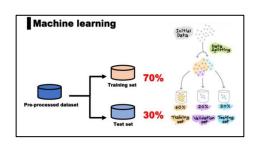


圖 9. 訓練、測試、分割

(time	設開	段件 :	ROE 5Y	EPS SY	毛利率5Y	現金融利率	28 .
3533	59x	76.9	16.59	409	26.47	2.95	1
4989	#H	26.05	14.61	3.07	13.05	5.57	1
1522	grat:	284	ma:	250	170	9.11	1
2012	5.89	17.05	0.12	104	402	3.42	1
3406	100	371.5	10.08	88.52	41.07	1.55	1
4961	未担	115.5	1966	6,67	25.47	3.86	1
3149	214	252	-031	457	-0.05	32/	-1
3136	992	160	12,01	2.19	1810	1.01	1
2434	MM	28.75	8150	1.72	28.21	6.0	1
1709	ha	27.45	7.17	1.14	16.75	4.36	-1
1460	212	7.24	-3,22	426	16.04	4.04	4
9942	.e/E	1085	13:49	532	3421	438	1
2915	単収立	010	28.85	96.51	45.40	7.83	1
4919	PE	120.0	13.28	107	39.45	2.08	1
2321	<b>元</b> 氏	34.15	-17.63	40	12.51	en	-1
6335	Ξ¢	10.3	+159.53	41	-0.016	6.0	-1
45/2	Ba	119.0	11,07	46	61.65	326	1
2085	800	213	871	0.2	2.22	3.16	-1

圖 10. 模型測試結果

# 参、系統架構

本系統共包含 4 個子系統。 分別為使用者 Web 介面子系統、網頁伺服 (Web Server)子系統、資料分析與繪圖子 系統、財報優良公司推薦子系統。(圖 11. 智慧財報網路平台開發系統架構圖)

『使用者 Web 介面子系統』顯示各家 上市公司的財報數值、線圖、各類股排序、 機器學習所列出的財報優良公司。

『網頁伺服(Web Server)子系統』根據使用者點選的功能,在資料庫抓取所需資料傳送至 Web 介面(前端)。

『資料分類計算、繪圖子系統』進行 爬蟲資料的分類計算,傳送到資料庫;取 得已分類計算的資料,將資料視覺化。

『財報優良公司推薦子系統』根據系統設置的正負樣本資料進行機器學習模型訓練,將訓練資料放置到測試,分出財報優良與不優良的公司,在Web介面顯示優良公司供使用者查看。

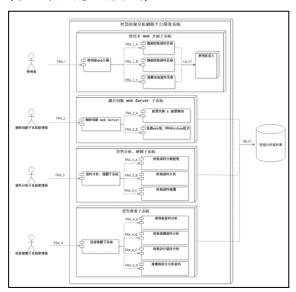


圖 11. 智慧財報網路平台開發系統架構圖

使用者 Web 介面子系統主要提供個股財報、類股財報、顯示財報優良公司, 當使用者登入後,會依照使用者所搜尋的 「股票代碼」與點選「類股」的分類來進 行財報的呈現;點選智慧財報分析公司頁 面,可查看經機器學習所分析的財報優良 公司。

本子系統共分四個部份,分別為登入元件、個股財報管理元件、類股財報管理元件、 財報優良公司元件。(圖 12. 使用者 Web 介面子系統架構圖)

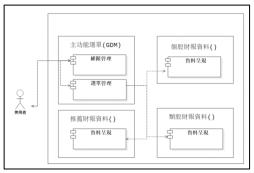


圖 12. 使用者 Web 介面子系統架構圖

網頁伺服 Web Server 介面子系統主要提供網頁伺服子系統管理員進行資料管理、個股財報管理、類股財報管理、財報優良公司管理,當網頁伺服子系統管理員登入子系統後,可以對客戶資料進行新增、刪除、修改和權限的調整,並依照客戶所點選的「股票代碼」、「產業別」、「智慧財報分析公司頁面」進行判讀,自動產生客戶相對應的資料表內容。

本子系統共分四個部份,分別為客戶 資料管理元件、個股財報管理元件、類股 財報管理元件、財報優良公司元件。(圖 13. 網頁伺服 Web Server 介面子系統)

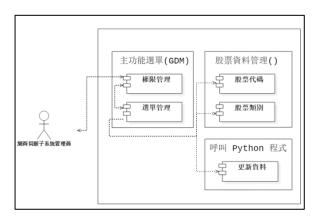


圖 13. 網頁伺服 Web Server 介面子系統 資料分類計算、繪圖子系統主要為財

報資料的分類計算,透過前置處理股票代碼、股票名稱與股票類別的爬取,並切割各季別的資料來加以計算,最終進行資料的合併,統整後存入資料庫,本作業系統會提供前端使用者所查詢的股票代碼、股票名稱來提供單季、近四季、年度…等相對應的資料與圖表線圖的呈現,也能依照股票類別來進行資料分類的呈現。

本子系統共分三個部份,分別為財報資料分類整理、財報資料計算、財報資料繪圖, (圖 14. 資料分類計算、繪圖子系統)除 了各元件的詳細說明外,本設計文件並將 對於各子系統中的功能與各參數進行詳細 敘述分別為 html 元件、Flask 元件、資料 庫元件…等。

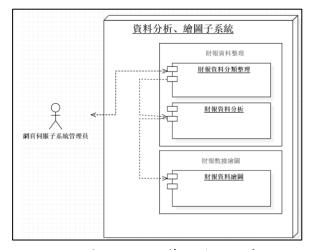


圖 14. 資料分類計算、繪圖子系統

量>=5)或-1(負樣本符合數量<=4)。利用 系統九項指標設置的標準,將各股票指標 進行 0 與 1 的分類,並將資料分割為訓練 7分與測試 3 分供模型訓練與測試,利用 模型進行訓練與測試並交互更換資料最後 得出股票推薦的結果,並將測試結果進行 呈現於 Web 介面。

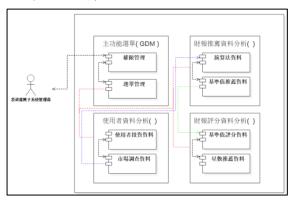


圖 15. 財報優良公司子系統

# 肆、系統實現

本組專題設計的智慧財報網路平台開發的實現做一驗證,下面分別為前面所介紹之功能呈現。本系統主要以 Python 程式語言編寫 Flask Web。

進入本系統後需要選擇登入或註冊帳號,讓使用者可輸入帳號與密碼或選擇註冊。主頁面使用 CSS 和 Bootstrap 排版(圖16. 登入畫面)

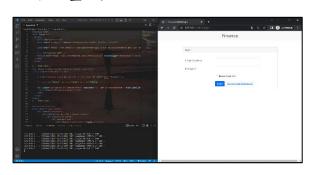


圖 16. 登入畫面

選擇註冊功能後便會跳轉到註冊網頁,透過在上方註冊讓使用者可以連結到 我們的網站並註冊,程式碼與頁面(圖 17. 註冊畫面)所示。

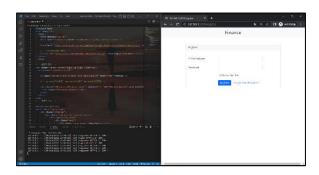


圖 17. 註冊畫面

登入後能依照使用者需求選擇想查詢的資料線圖,包含了單季、近四季、年度數值與線圖,此功能使用了 JQuery UI AutoComplete 讓使用者搜尋股票時,自動從資料庫搜尋是否有符合的股票代碼,減少使用者填錯代號的可能性,線圖的部分使用 Pycharts 套件中的線圖功能(圖 18.線圖畫面)。



圖 19. 線圖畫面

另外也有讓使用者查詢各類股的數值 資料,我們透過 Bootstrap 中的 datasort 排序的方式,將數值由大至小往下排,方 便使用者觀看該類股排行(圖 20. 類股排 序畫面)。



圖 20. 類股排序畫面

使用機器學習模型的正負樣本訓練資

料(圖 21.正負樣本),利用支持向量機 (Support Vector Machine)與決策樹 (Decision Tree)進行訓練與測試,訓練正確率高達 98%,測試正確率為 96%,網頁平台的部分,也提供表格的呈現,提供更清晰的公司好壞分析結果(圖 22.分析結果)。

Α	В		D	E	F	0	H			K	L	М	N		P	Q
代碼	股票	ROE_5Y	EPS_5Y	毛利率_9	現金設和	成交債	量監持部	現金雅利	ROE_SYL	EPS_SYF	毛利率	現金殖利	設備』	量監持服	Sum	判斷
1101	台泥	10.26	3,37	25.68	2.34	34.25	8.38	6.83	1	1	1	1	1	. 0		5
1102	EEE	8.44	3,36	22.76	2.77	40	27.37	6.93	1	1	1	1	1	1		6
1103	嘉泥	4.67	1.46	6.77	0.91	17.05	17.15	5.34	0	1	(	) 1	1	. 1		4
1104	现尼	7.57	2.02	13.3	1.04	20.45	24.16	5.09	1	- 1		1	1	- 1		5
1108	幸福	1.89	0.23	8.91	0.58	10.35	3.95	5.6	0	1	(	1	1	. 0		3
1109	信大	12.95	1.98	27.76	1	15,65	36.26	6.39	1	1	1	1	1	1		6
1110	東泥	0.89	0.1	3.24	0.12	17,65	34,67	0.68	0	1	(	0	1	. 1		3
1203	味王	12.19	1.84	31.49	1.1	31.8	28.84	3.46	1	1	1	. 0	1	1		5
1210	大成	13	3.19	14.12	1.98	44,75	12.98	4.42	1	1	(	0	1	1		4
1213	大飲	4.72	-0.5	7.92	0	7.3	12.91	0	0	0	(	0	1	. 1		2
1215	卜蜂	19.9	5.05	15.18	3.5	78.6	10	4.45	1	1	(	0	1	1		4
1216	統一	19.81	3,96	33.79	3.18	65.7	10.05	4.84	1	1	1		1	1		5
1217	爱之味	2.02	0.23	31.9	0.06	9.4	10.24	0.64	0	1	1	. 0	1	1		4
1218	泰山	8.55	1.04	18.17	0.9	34.85	35.9	2.58	1	1	(	0	1	1		4
1219	福舞	3.5	0.69	8.64	0.58	19.45	6.9	2.98	0	1	(	0	1	0		2
1220	台発	5.9	0.73	13.58	0.69	14.4	21.73	4.79	1	1	(	0	1	. 1		4
1225	基度油	10.58	1.47	9.72	1.34	0	22,84	0	1	1	(	0	1	1		4
1227	佳格	18.51	3.18	29.28	2.31	38.4	3.2	6.02	1	1	1	1	1	. 0		5
1229	聯展	10.23	2,55	20.85	1.7	49,85	23.62	3.41	1	1	1	. 0	1	1		5
1231	聯幕会	15.7	3.12	21.38	1.58	70.1	21.79	2.25	1	1			1	1		5
1233	天仁	9.69			1.47	33.2	11.07	4.43	1	1	1	0	1	1		5
1234	題松	4.6	2.03	26.28	1.66	33.3	12.06	4.98	0	1	1	0	1	1		4
1235	興泰	5.08			0.61	62.6		0.97	1	1	1	. 0	1	1		5
1236	宏亞	1.18	0.2	27.23	0.32	21.6	24.47	1.48	0	1	1	. 0	1	1		4
	鲜活果汁					230		3.69		1	1	. 0	0	0		3
1301	台塑	11.75	6.15	15.39	5.3	85.1	14.93	6.23	1	1	(	1	0	1 1		4

圖 21. 正負樣本

代碼	股票	ROE_5Y	EPS_5Y	毛利率_5	成交價	董監持股	現金殖利	答案
2385	群光	18	6.53	17.1	84	10.47	6.13	1
2450	神腦	10.41	2.44	11.92	30	33.53	6.37	1
4807	日成-KY	17.13	2.64	25.42	19.5	43	12.46	1
9931	欣高	10.25	1.84	17.79	38.5	24.83	2.96	1
2104	國際中核	7.38	2.46	24.7	19.65	0	3.87	1
2331	精英	-0.7	-0.21	10.74	25.2	32.5	9.52	1
1309	台達化	13.65	1.86	9	20.65	39.37	4.55	1
3653	健策	9.87	4.57	26.81	418.5	25.89	1.08	1
6672	騰輝電子	39.5	6.59	25.14	62.1	14.13	6.17	1
2379	瑞昱	20.91	10.38	43.51	259.5	6.21	4.97	1
2423	固緯	12.36	1.92	49.59	25.1	8.94	5.98	1
1615	大山	11.81	1.75	12.32	28.3	15.43	4.56	1
2327	國巨	30.19	29.34	39.11	287	7.57	6.48	1
2489	瑞軒	2.25	0.36	10.13	10	11.73	12	1
3005	神基	14.76	3.79	28.05	43	34.13	7.21	1

圖 22. 分析結果

## 伍、結論

本專題在製作專案的過程中結合網路 爬蟲、資料處理計算、資料繪圖、機器學 習相對應的技術,並透過網頁設計與資料 庫結合來呈現給使用者操作,除了像一般 財報網站給予公司財報資料外,我們另外 提供給使用者機器學習測試後所分析出的 財報優良公司,我們也會列出利用不同機 器學習模型,測試出的結果,讓使用者可 以交叉比對,給予使用者更多的參考價 值。

本專題為了確保基本盤資訊的準確度,除了參考財報狗、玩股網、公開觀測站…等相關財報網站,也加入了四項財報指標進行正負樣本分類,利用 Data Science資料科學的分析技術,並搭配機器學習方法-支持向量機(Support Vector Machine)與決策樹(Decision Tree)進行模型訓練與測試,讓使用者能查看財報優良的公司。

## 參考文獻

- [1]. 財報狗團隊,財報狗資訊股份有限公司,財報狗, https://statementdog.com, 2022。
- [2]. 玩股網,<u>https://www.wantgoo.com</u>。
- [3]. 公開資訊觀測站,

https://mops.twse.com.tw/mops/web/inde <u>x</u> °

[4]. 10 程式中, [Day 11] 核模型 - 支持向量機 (SVM), iT 邦幫忙, 2021

年,<u>https://ithelp.ithome.com.tw/articles/10</u> 270447