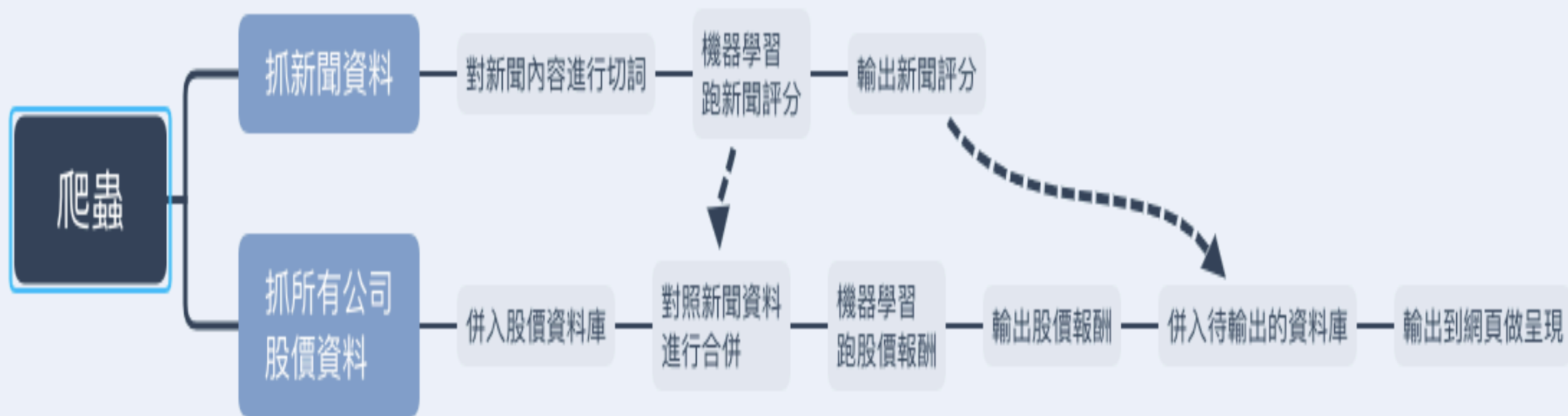


第二次進度報告

玉山證券-重大事件訊息揭露，對個股影響強度預測模型

組員:劉品妤，王昱達，楊廣元，呂明諺

專案流程圖





看門狗資料簡介與切詞處理



資料與問題定義

2019年看門狗資訊事件整理(全年) - Excel

品妤 劉

檔案 常用 插入 頁面配置 公式 資料 校閱 檢視 開發人員 說明 告訴我您想做什麼 共用

貼上 剪貼簿 字型 對齊方式 數值 樣式 儲存格 編輯

A1 個股代號

	A	B	C	D	E	F	G	H	I	J	K	L
1	個股代號	公司簡稱	事件日	TCRI(年/月)	事件強度	大事件類別	小事件類別	事件內容				
2	1218	泰山	20190101	6(2018/09)	0	M_經營層	MT06_高管異動	發言人林俐婉內部調動，由江巍峰接任。。				
3	1503	士電	20190101	4(2018/09)	0	M_經營層	MT06_高管異動	內部稽核主管林志強內部調動，由莊文清接任。。				
4	1504	東元	20190101	4(2018/09)	0	M_經營層	MT06_高管異動	會計主管藍俊雄內部調動，由林鴻名接任。。				
5	1709	和益	20190101	5(2018/09)	0	M_經營層	MT06_高管異動	內部稽核主管游本詮內部調動，由曾筱茜接任。。				
6	1721	三晃	20190101	7(2018/09)	0	M_經營層	MT06_高管異動	財務經理洪廷宜內部調動，由王婷渝接任。。				
7	1817	凱撒衛	20190101	6(2018/09)	0	M_經營層	MT06_高管異動	研發主管吳政峰內部調動，由朱清立接任。。				
8	2064	晉椿	20190101	7(2018/09)	-1	M_經營層	MT06_高管異動	總經理高進義離職，由陳譽接任。。發言人高進義離職				
9	2207	和泰車	20190101	4(2018/09)	0	M_經營層	MT02_董監異動	改派1董。董事大野勝仁(豐田自動車代表)卸任。董事長				
10	2330	台積電	20190101	1(2018/09)	-1	M_經營層	MT02_董監異動	辭任1董。獨立董事湯馬斯?延吉布斯卸任。				
11	2357	華碩	20190101	2(2018/09)	0	M_經營層	MT06_高管異動	總經理沈振來內部調動，由胡書賓接任。。				
12	2377	微星	20190101	3(2018/09)	0	M_經營層	MT06_高管異動	總經理徐祥內部調動，由江勝昌接任。。				
13	2442	新美齊	20190101	7(2018/09)	0	M_經營層	MO04_經營權轉讓疑慮	新美齊澄清報載提及本公司停業及隱射屬大同集團上市				
14	2724	富驛-KY	20190101	D(2018/09)	0	M_經營層	MT02_董監異動	改派1董。董事周威良(Furama Hotel International Managem				
15	2750	桃禧	20190101		0	M_經營層	MT06_高管異動	會計主管林慧茹內部調動，由楊崑岳接任。。				
16	2852	第一保	20190101		0	M_經營層	MT06_高管異動	財務經理李易致內部調動，由施冬森接任。。				

全年

100%

資料預處理(斷詞)

- 詞是最小有意義且可以自由使用的語言單位
- 任何語言處理的系統都必須先能分辨文本中的詞才能進行進一步的處理
- 本次報告須做處理的部分為新聞的內容
- 分別使用2種不同的斷詞系統做新聞斷詞

JIEBA

- Jieba 這個中文斷詞程式是由中國開發者所開發
- 可同時支援簡體與繁體的斷詞

CKIPTAGGER

- 研究的中研院CKIP Lab中文詞知識庫小組開發之中文斷詞工具
- CkipTagger表現遠高於中國的結巴，中研院在中文斷詞準確度可達到97.49%，相較之下，中國的結巴只有90.51%

資料預處理(斷詞 JIEBA V.S. CKIPTAGGER)

TOOL	(WS)PREC	(WS)REC	(WS)FI
CkipTagger	97.49%	97.17%	97.33%
jieba	90.51%	89.10%	89.80%

資料預處理(斷詞 JIEBA V.S. CKIPTAGGER)



fintech - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明

```
jieba.add_word("發言人")
jieba.add_word("林俐婉")
jieba.add_word("江巍峰")
jieba.add_word("林志強")
jieba.add_word("莊文清")
jieba.add_word("徐祥")
jieba.add_word("江勝昌")
jieba.add_word("大同")
jieba.add_word("集團")
jieba.add_word("壁紙")
# jieba.suggest_freq((' ', '經理'), True)
jieba.add_word("內部")
jieba.add_word("調動")
jieba.add_word("會計")
jieba.add_word("財務")
jieba.add_word("經理")
jieba.add_word("研發")
jieba.add_word("主管")
jieba.add_word("洪廷宜")
jieba.add_word("王婷渝")
jieba.add_word("吳政峰")
jieba.add_word("朱清立")
jieba.add_word("調動")
jieba.add_word("新美齊")
jieba.add_word("隱射")
```

In [17]: word_sentence_list = ws(sentence_list)

```
# pos_sentence_list = pos(word_sentence_list)
```

```
# entity_sentence_list = ner(word_sentence_list, pos_sentence_list)
```

In [18]: content_tokenized=[]

```
In [19]: for i in word_sentence_list:
          content_tokenized.append(' '.join(i))
```

In [20]: content_tokenized=pd.DataFrame(content_tokenized)

```
In [21]: content_tokenized.columns=["content_tokenized"]
          content_tokenized
```

Out[21]:

	content_tokenized
0	發言人 林俐婉 內部 調動 由 江巍峰 接任
1	內部 稽核 主管 林志強 內部 調動 由 莊文清 接任
2	會計 主管 藍俊雄 內部 調動 由 林鴻名 接任
3	內部 稽核 主管 游本詮 內部 調動 由 曾筱茜 接任
4	財務 經理 洪廷宜 內部 調動 由 王婷渝 接任
5	研發 主管 吳政峰 內部 調動 由 朱清立 接任
6	總經理 高進義 離職 由 陳譽 接任 發言人 高進義 離職 由 陳譽 接任
7	改派 1 董 董事 大野勝仁 豐田 自動車 代表 卸任 董事長 沼一生 豐田 自動車 代表 新任



新聞資料爬蟲



新聞資料爬蟲

- 本組採用 pandas 套件中的 read_html 功能，對公開資訊觀測站的重大訊息主旨做爬蟲
- 設定成每10秒爬一次最新消息，將新增的新聞訊息加入資料庫，並丟入模型跑重要性評分與預期股價波動

3:00~14:00進行系統維修，期間服務短暫中斷，不便之處敬請見諒

English

請輸入公司代號、簡稱，或報表關鍵字 搜尋

常用 營收 除權息 電子書 法說會 庫藏股 董監持股 獨立董事 董監酬金 ETF TDR

常用報表 基本資料 彙總報表 股東會及股利 公司治理 財務報表 重大訊息與公告 營運概況 投資專區 認購(售)權證 債券

重大訊息與公告

即時重大訊息

重大訊息綜合查詢

臺灣存託憑證收盤價彙總表

法說會

公告查詢

券商對媒體轉載之澄清或說明

即時重大訊息

列印網頁 開新視窗 問題回報

市場別：全體公司

全體公司 上市公司 上櫃公司 興櫃公司 公開發行公司

公司代號	公司簡稱	發言日期	發言時間	主旨	詳細資料
4133	亞諾法	109/06/11	09:56:40	本公司簽訂COVID-19治療性抗體人類化合作協議	詳細資料
6679	鈺太	109/06/11	09:35:47	澄清媒體報導	詳細資料
3625	西勝	109/06/11	07:37:58	公告本公司成立第一屆審計委員會	詳細資料
3625	西勝	109/06/11	07:25:59	公告本公司109年度股東常會改選董監事當選名單	詳細資料
3073	天方能源	109/06/11	07:00:01	公告本公司名稱由「凱柏實業股份有限公司」更名為「天方能源科技股份有限公司」	詳細資料



機器學習建模



模型建立

利用長短期記憶模型（LSTM）建立：

1. 大事件類別分類器
2. 小事件類別分類器
3. 事件強度分類器
4. 股價異常報酬分類器

大事件分類器

利用新聞中的文字資料，將新聞分類為以下五個大事件類別：

- 'A_會計/財報分析'
- 'F_市場交易'
- 'I_產業前景'
- 'M_經營層'
- 'R_危機'

大事件分類器：資料分割

- 所有資料的64%作為訓練資料 (training set)
- 所有資料的16%作為驗證集 (validation set)
- 所有資料的20%作為測試集 (testing set)

大事件分類器：處理“不平衡資料” (IMBALANCED DATA)

- 由於大事件類別的分布相當不平衡，所以我們使用了以下兩種方法來處理資料不平衡的問題
- 1. 使用Oversampling
- 2. 調整損失函數 (loss function) 的權重

大事件分類器：模型架構

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, None, 128)	1280000
lstm_2 (LSTM)	(None, 16)	9280
dense_2 (Dense)	(None, 5)	85
Total params: 1,289,365		
Trainable params: 1,289,365		
Non-trainable params: 0		

大事件分類器：模型架構

- Embedding layer: 用來進行詞嵌入 (word embedding)
- LSTM layer: 長短期記憶模型
- Dense layer: 作為此模型的output layer

大事件類別分類器：模型表現（在驗證集上）

accuracy: 0.971

	'A_會計/財報分析'	'F_市場交易'	'I_產業前景'	'M_經營層'	<u>'R_危機'</u>
precision:	[0.672	0.995	0.979	0.959	0.888]
recall:	[0.741	0.974	0.982	0.96	<u>0.879]</u>
F1 score:	[0.705	0.984	0.980	0.960	0.883]

（為了維持易讀性，將所有數字取四捨五入到小數點後三位數。）

小事件分類器

利用新聞中的文字資料，將新聞分類為以下15小事件類別：

'AF05_財務警示' 'AI01_延遲公告' 'FS02_股價暴跌或異常' 'FS03_其他市場交易議題'
'IP01_成本/產能變動或資本支出' 'IS01_營收變動或客戶/商品/通路策略' 'MT02_董監異動'
'MT06_高管異動' 'RB01_TCRI負向觀察' 'RB02_TCRI降等'
'危機_其他' '市場交易_其他' '會計/財報分析_其他' '產業前景_其他' '經營層_其他'

小事件分類器：資料分割

- 所有資料的64%作為訓練資料 (training set)
- 所有資料的16%作為驗證集 (validation set)
- 所有資料的20%作為測試集 (testing set)

小事件分類器：處理“不平衡資料” (IMBALANCED DATA)

- 由於小事件類別的分布相當不平衡，所以我們使用了以下兩種方法來處理資料不平衡的問題
- 1. 使用Oversampling
- 2. 調整損失函數 (loss function) 的權重

小事件分類器：模型架構

Layer (type)	Output Shape	Param #
embedding_6 (Embedding)	(None, None, 128)	1280000
lstm_6 (LSTM)	(None, 64)	49408
dense_15 (Dense)	(None, 32)	2080
dense_16 (Dense)	(None, 32)	1056
dense_17 (Dense)	(None, 15)	495
Total params: 1,333,039		
Trainable params: 1,333,039		
Non-trainable params: 0		

小事件分類器：模型架構

- Embedding layer: 用來進行詞嵌入 (word embedding)
- LSTM layer: 長短期記憶模型
- Dense layers (3 dense layers): 進行小事件類別的分類

小事件類別分類器：模型表現（在驗證集上）

accuracy: 0.905

['AF05_財務警示' 'AI01_延遲公告' 'FS02_股價暴跌或異常' 'FS03_其他市場交易議題' 'IP01_成本/產能變動或資本支出' 'IS01_營收變動或客戶/商品/通路策略' 'MT02_董監異動' 'MT06_高管異動' 'RB01_TCRI負向觀察' 'RB02_TCRI降等' '危機_其他' '市場交易_其他' '會計/財報分析_其他' '產業前景_其他' '經營層_其他']

precision: [0.128 1. 1. 0.8 0.797 0.943 1. 1. 0.806 0.888 0.941 0.8 0.7 0.759 0.907]

recall: [0.451 0.375 0.999 0.333 0.701 0.945 0.995 0.995 0.781 0.8 0.592 0.470 0.491 0.839 0.870]

F1 score: [0.191 0.545 0.999 0.470 0.746 0.944 0.997 0.997 0.793 0.842 0.727 0.592 0.577 0.797 0.888]

（為了維持易讀性，將所有數字取四捨五入到小數點後三位數。）

事件強度分類器

利用新聞中的文字資料，將新聞分類為以下七個事件強度：

- -3, -2, -1, 0, 1, 2, 3

事件強度分類器：資料分割

- 所有資料的64%作為訓練資料 (training set)
- 所有資料的16%作為驗證集 (validation set)
- 所有資料的20%作為測試集 (testing set)

事件強度分類器：處理“不平衡資料” (IMBALANCED DATA)

由於事件強度的分布相當不平衡（極端事件：-3, -2, +2, +3 出現的頻率相對較少），所以我們使用了以下兩種方法來處理資料不平衡的問題

- 1. 使用Oversampling
- 2. 調整損失函數（loss function）的權重

事件強度分類器：模型架構

Layer (type)	Output Shape	Param #
embedding_3 (Embedding)	(None, None, 128)	1280000
lstm_3 (LSTM)	(None, 16)	9280
dense_3 (Dense)	(None, 7)	119
Total params: 1,289,399		
Trainable params: 1,289,399		
Non-trainable params: 0		

事件強度分類器：模型架構

- Embedding layer: 用來進行詞嵌入 (word embedding)
- LSTM layer: 長短期記憶模型
- Dense layer: 作為此模型的output layer

事件強度分類器：模型表現（在驗證集上）

- accuracy: 0.8779956427015251

■ 事件強度類別	[-3,	-2,	-1,	0,	1,	2,	3]
■ precision:	[0.333	0.629	0.957	0.829	0.789	0.857	1.]
■ recall:	[0.710	0.542	0.941	0.843	0.770	0.5	1.]
■ F1 score:	[0.454	0.582	0.949	0.836	0.779	0.632	1.]

（為了維持易讀性，將所以數字取四捨五入到小數點後三位數。）



股價預測



股價預測——事件研究法

何謂事件研究？

事件研究法(Event Study) 為研究結果之驗證方法，其起源於1960年代 Ball and Brown，及 Fama, Fisher, Jensen and Roll (沈中華、李建然，2000)，為近代會計及財務領域實證研究所廣泛運用之研究設計之一。

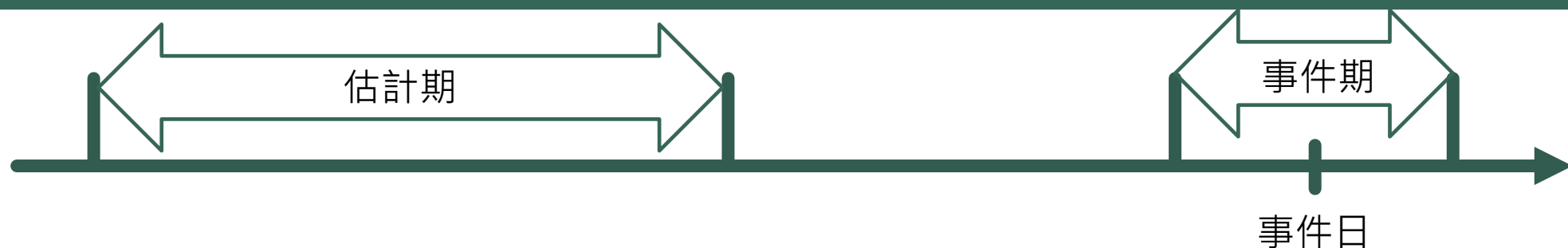
事件研究的目的？

事件研究法(Event Study) 主要目的在於**利用統計方法檢定異常報酬狀況**，藉以明瞭特定事件是否**對公司股價造成影響**，並可以了解股價的波動與該事件是否相關。

事件研究流程

1. 決定事件與事件日
2. 估計異常報酬率
3. 檢定異常報酬率
4. 分析結果及解釋

股價預測——事件日、事件期、估計期



如何設定事件期、估計期長度

- 事件影響的區間應包括在事件期之內，如新聞發布之日。通常事件期間比發生日期（事件日）更寬廣一些，包括事件發生前後的一段時間。因為**事件發生後一段時間**的資訊能顯示**應變數（如盈利、股價）變化**的情況；而考察**事件發生後一段時間**的股價則有利於捕捉**事件前徵兆**與**事前洩漏資訊**所造成的影響。
- 估計期間或稱**估計窗口（estimation window）**的目的，是利用該期間的數據去估算在事件未出現情況下應變數之值，即**預期報酬率**。將預期報酬率與事件期間應變數變異後（即實際報酬率）相比較，變得出事件所帶來的異常報酬率。
- 一般而言，估計期選取要比事件期間長，本組採用年（250 個交易日）、季（60 個交易日）、月（20 個交易日）三個區間去估計異常報酬，而事件期則是事件日前後一天（明日收盤價 - 昨日收盤價）

股價預測——異常報酬計算結果

個股代號	公司簡稱	事件日	TCRI(年/月)	事件強度	收盤價	明日收盤價 / 昨日收盤價	明日收盤價 / 昨日收盤價 - 1	兩日報酬率(%)	前5 ~ 245日平均兩日報酬率(%)，年平均正常報酬(%)	前5 ~ 65日平均兩日報酬率(%)，季平均正常報酬(%)	前5 ~ 25日平均兩日報酬率(%)，月平均正常報酬(%)	前5 ~ 245日(年平均)兩日異常報酬(%)	前5 ~ 65日(季平均)兩日異常報酬(%)	前5 ~ 25日(月平均)兩日異常報酬(%)	大事件類別	小事件類別	事件內容	切詞詞數	切詞結果
1218	泰山	20190101	6(2018/09)	0	19	0.972973	-0.02703	-2.7027	0.072008802	-0.126391428	0.91151213	-2.774711504	-2.576311275	-3.614214833	M_經營層	MT06_高	發言人林	19	['發言人', '林俐婉', '內部', '調動', '。', '。']
1503	士電	20190101	4(2018/09)	0	40.95	0.99754	-0.00246	-0.246	-0.000782104	-0.459070858	-0.042079001	-0.245220356	0.213068398	-0.203923459	M_經營層	MT06_高	內部稽核	22	['內部', '稽核', '主管', '林志強', '內部', '調動', '。', '。']
1504	東元	20190101	4(2018/09)	0	17.45	0.994302	-0.0057	-0.5698	-0.373847067	-0.699126047	0.436006766	-0.195953503	0.129325477	-1.005807335	M_經營層	MT06_高	會計主管	20	['會計', '主管', '藍俊雄', '內部', '調動', '。', '。']
1709	和益	20190101	5(2018/09)	0	14.8	1.003367	0.003367	0.3367	-0.097104514	-0.364537588	-0.160869687	0.433804851	0.701237924	0.497570024	M_經營層	MT06_高	內部稽核	22	['內部', '稽核', '主管', '游本詮', '內部', '調動', '。', '。']
1721	三晃	20190101	7(2018/09)	0	10.05	1	0	0	-0.275157057	-0.026064579	0.557660856	0.275157057	0.026064579	-0.557660856	M_經營層	MT06_高	財務經理	20	['財務', '經理', '洪廷宜', '內部', '調動', '。', '。']
1817	凱撒衛	20190101	6(2018/09)	0	36.95	0.989276	-0.01072	-1.07239	-0.15792161	-0.081807145	0.3658829	-0.914464449	-0.990578914	-1.438268959	M_經營層	MT06_高	研發主管	20	['研發', '主管', '吳政峰', '內部', '調動', '。', '。']
2064	晉椿	20190101	7(2018/09)	-1	15.55	1.01634	0.01634	1.633987	-0.069460762	-0.219389167	-0.009464759	1.70344769	1.853376095	1.643451687	M_經營層	MT06_高	總經理高	32	['總經理', '高進義', '離職', '。', '。', '由', '陳']
2207	和泰車	20190101	4(2018/09)	0	255.5	0.988	-0.012	-1.2	-0.262075048	-0.048626627	1.179844775	-0.937924952	-1.151373373	-2.379844775	M_經營層	MT02_董	改派1董	41	['改派', '1', '董', '。', '董事', '大野勝仁', '。']
2330	台積電	20190101	1(2018/09)	-1	225.5	0.984305	-0.0157	-1.56951	-0.024201137	-0.551439256	0.027877375	-1.545305589	-1.01806747	-1.597384101	M_經營層	MT02_董	辭任1董	20	['辭任', '1', '董', '。', '獨立', '董事', '湯馬']
2357	華碩	20190101	2(2018/09)	0	201.5	1.0175	0.0175	1.75	-0.240915578	-0.822803805	-0.693052721	1.990915578	2.572803805	2.443052721	M_經營層	MT06_高	總經理沈	19	['總經理', '沈振來', '內部', '調動', '。', '。']
2377	微星	20190101	3(2018/09)	0	76.4	1.005222	0.005222	0.522193	-0.026185798	-0.34696121	0.924262083	0.548379009	0.869154422	-0.402068872	M_經營層	MT06_高	總經理徐	18	['總經理', '徐祥', '內部', '調動', '。', '。', '由']
2442	新美齊	20190101	7(2018/09)	0	12	1.0125	0.0125	1.25	0.460998352	-0.435840484	-0.448641991	0.789001648	1.685840484	1.698641991	M_經營層	MO04_經	新美齊潛	536	['新美齊', '澄清', '報載', '提及', '本', '公']
2724	富驛-KY	20190101	D(2018/09)	0	2.98	0.993333	-0.00667	-0.66667	0	-0.156359079	-2.322596508	0	-0.510307588	1.655929841	M_經營層	MT02_董	改派1董	113	['改派', '1', '董', '。', '董事', '周威良', '。']
2750	桃禧	20190101		0	18.11	1	0	0	0	0.224634033	-0.510253045	0	-0.224634033	0.510253045	M_經營層	MT06_高	會計主管	20	['會計', '主管', '林慧茹', '內部', '調動', '。', '。']
2852	第一保	20190101		0	14.7	0.996575	-0.00342	-0.34247	0.019577627	-0.203529087	0.005481137	-0.36204338	-0.138936667	-0.34794689	M_經營層	MT06_高	財務經理	20	['財務', '經理', '李易致', '內部', '調動', '。', '。']
2888	新光金	20190101		0	8.97	0.962596	-0.0374	-3.74037	-0.088140176	-0.84149414	-1.022169478	-3.652233861	-2.898879898	-2.71820456	M_經營層	MT06_高	會計主管	20	['會計', '主管', '施貽桓', '內部', '調動', '。', '。']
3004	豐達科	20190101	5(2018/09)	0	59.9	1.01005	0.01005	1.005025	0.168964124	0.292028166	0.401267721	0.836061002	0.71299696	0.603757405	M_經營層	MT06_高	總經理邱	19	['總經理', '邱智科', '內部', '調動', '。', '。']
3167	大量	20190101	5(2018/09)	0	39.95	0.997481	-0.00252	-0.25189	-0.196109258	-0.368037992	-0.637395927	-0.055779911	0.116148823	0.385506758	M_經營層	MT06_高	研發主管	20	['研發', '主管', '宋漢訓', '內部', '調動', '。', '。']
4104	佳醫	20190101	5(2018/09)	0	46	1.001088	0.001088	0.108814	0.023426418	-0.099172149	0.155574448	0.085387511	0.207986078	-0.04676052	M_經營層	MT06_高	總經理高	18	['總經理', '高省', '內部', '調動', '。', '。', '由']
4138	曜亞	20190101	6(2018/09)	0	46.6	1.008621	0.008621	0.862069	-0.051387204	-0.267159014	-0.025072178	0.91345617	1.12922798	0.887141143	M_經營層	MT06_高	總經理傅	19	['總經理', '傅若軒', '內部', '調動', '。', '。']
4168	聯聯	20190101	7(2018/09)	0	23.1	0.995633	-0.00437	-0.43668	0.118934797	-0.542141067	-0.305876173	-0.55561602	0.105459844	-0.130805049	M_經營層	MT06_高	總經理張	19	['總經理', '張東玄', '內部', '調動', '。', '。']
4402	福大	20190101	D(2018/09)	0	4.31	0.855649	-0.14435	-14.4351	0	-2.458430891	0.312636652	0	-11.97671555	-14.7477831	M_經營層	MT06_高	會計主管	20	['會計', '主管', '莊清揚', '內部', '調動', '。', '。']
4535	至興	20190101	5(2018/09)	0	31.15	0.998395	-0.00161	-0.16051	-0.254834581	-0.621912893	-0.522904249	0.094320937	0.461399249	0.362390605	M_經營層	MT06_高	研發主管	20	['研發', '主管', '蕭永隆', '內部', '調動', '。', '。']
4726	永昕	20190101	7(2018/09)	-1	25.8	1.007663	0.007663	0.766284	-0.033462067	0.343059567	1.07502444	0.799745592	0.423223958	-0.308740915	M_經營層	MT06_高	總經理潘	74	['總經理', '潘國蘭', '離職', '。', '。', '同年', '12']
4987	科誠	20190101	5(2018/09)	0	54.7	0.998172	-0.00183	-0.18282	-0.083506725	-0.323015309	0.202024762	-0.099308632	0.140199952	-0.384840119	M_經營層	MT06_高	總經理顧	19	['總經理', '戴奉義', '內部', '調動', '。', '。']
6201	亞弘電	20190101	5(2018/09)	0	29.55	1.008489	0.008489	0.848896	0.019842577	-0.071658977	-0.006430991	0.829053858	0.920555411	0.855327426	M_經營層	MT06_高	總經理許	38	['總經理', '許文鼎', '內部', '調動', '。', '。']
6221	晉泰	20190101	6(2018/09)	0	28.05	0.994652	-0.00535	-0.53476	0.166185771	0.054287866	0.756790184	-0.700945129	-0.589047224	-1.291549542	M_經營層	MT06_高	總經理林	19	['總經理', '林澤鉅', '內部', '調動', '。', '。']
6251	定穎	20190101	6(2018/09)	0	8.9	1.006772	0.006772	0.677201	-0.181733472	0.154118241	0.396099809	0.858934375	0.523082662	0.281101094	M_經營層	MT06_高	會計主管	20	['會計', '主管', '江玲聲', '內部', '調動', '。', '。']
6407	相互	20190101	8(2018/06)	0	23.19	1.007343	0.007343	0.734341	-0.023136591	-0.47067266	0.1541108	0.757477843	1.205013913	0.580230453	M_經營層	MT06_高	總經理林	19	['總經理', '林錦鴻', '內部', '調動', '。', '。']
6409	旭華	20190101	3(2018/09)	-1	537	0.974265	-0.02574	-2.57353	0.027403212	-0.079367293	0.006785625	-2.600932624	-2.494162118	-2.580315037	M_經營層	MT02_董	辭任1董	14	['辭任', '1', '董', '。', '獨立', '董事', '陳一']
6599	普達系統	20190101	6(2018/06)	0	17.5	0.988764	-0.01124	-1.1236	-0.482250803	-0.876615412	0.797398193	-0.641344703	-0.246980094	-1.920993699	M_經營層	MT06_高	研發主管	20	['研發', '主管', '莊漢祺', '內部', '調動', '。', '。']
8176	智捷	20190101	7(2018/09)	0	10.8	0.99537	-0.00463	-0.46296	-0.404886238	0.108786105	0.594841115	-0.058076725	-0.571749068	-1.057804078	M_經營層	MT06_高	總經理謝	19	['總經理', '謝金生', '內部', '調動', '。', '。']
8458	影一	20190101		0	4.15	1	0	0	-0.549122565	-1.502617036	-1.5886596	0.549122565	1.502617036	1.5886596	M_經營層	MT06_高	內部稽核	62	['內部', '稽核', '主管', '鄭健達', '內部', '調動', '。', '。']

股價異常報酬分類器

- 預測股價的異常報酬為正值或負值（請見下一個章節「股價預測」的投影片）
- 將「預測股價異常報酬」的問題視為二分類問題（正值 or 負值）

股價異常報酬分類器：資料分割

- 所有資料的64%作為訓練資料 (training set)
- 所有資料的16%作為驗證集 (validation set)
- 所有資料的20%作為測試集 (testing set)

股價異常報酬分類器：模型架構

Layer (type)	Output Shape	Param #
embedding_4 (Embedding)	(None, None, 128)	1280000
lstm_4 (LSTM)	(None, 64)	49408
dense_13 (Dense)	(None, 32)	2080
dense_14 (Dense)	(None, 32)	1056
dense_15 (Dense)	(None, 32)	1056
dense_16 (Dense)	(None, 1)	33
Total params: 1,333,633		
Trainable params: 1,333,633		
Non-trainable params: 0		

股價異常報酬分類器：模型架構

- Embedding layer: 用來進行詞嵌入 (word embedding)
- LSTM layer: 長短期記憶模型
- Dense layers (4 dense layers): 進行股價異常報酬的分類

股價異常報酬分類器：模型表現（在驗證集上）

- accuracy: 0.596
- 事件強度類別 [負, 正]
- precision: [0.582 0.619]
- recall: [0.716 0.474]
- F1 score: [0.642 0.537]

（為了維持易讀性，將所有數字取四捨五入到小數點後三位數。）

股價預測——結果分析

模型結果

- 若將評分與異常報酬做簡單迴歸分析，R平方只有0.1，代表平均來說，新聞事件的變動只能解釋10%的股價變動，解釋力不足
- 若只預測上漲與下跌（也就是只分兩個類別），Accuracy Rate 只有 0.596，還有進步空間

結果解釋

- 新聞事件對股價影響的反應時間極短，可能事件前後幾小時內股價就已經反應完畢，而本組採用的兩日區間過長，造成解釋力不足的結果（但受限於沒有 Intra Day 的股價資料，只能這麼做）
- 不重要的新聞事件數占資料絕大部分，而這些新聞對股價影響力不大，卻又大量使用這些資料做機器學習，將評分絕對值高的事件的影響力給稀釋掉



資料庫建立與網頁呈現



資料庫建立

在將資料輸入上傳到網頁前，
我們會將：

1. 股票代碼
2. 公司名稱
3. 發生時間
4. 異常報酬
5. 事件強度評分
6. 新聞內容
7. 新聞切詞結果

等資料存放在資料庫中，未來公司需要做使用即可直接做存取



1. 可使用 gspread 套件與 python 做連接
2. 有容量限制，不適合放大量數據
3. 每 100 秒有資料存取量限制，不適合作為存取頻繁的資料庫
4. 免費試用期一年，之後要收費



1. 可使用 pymysql 套件與 python 做連接
2. 無容量限制，適合放大量數據
3. 無資料存取量限制，適合作為存取頻繁的資料庫
4. 開源軟體，免費使用

- 採用 AWS 與 WordPress 架構個股新聞評分系統網站，並利用 wordpress-xmlrpc 套件完成自動發文的功能
- 我們預期將使用爬蟲爬取最新新聞內容後，經過我們的評分與股價預測系統模型預測，將重要的新聞（評分絕對值大於2的新聞）呈現在我們的網站中

結果呈現



首頁



20190101 新美齊

ON 2020-05-01 / 由 P314008DSA1 撰寫 / IN 2442 新美齊 / 編輯

事件強度：0
預期漲跌幅：0.2%

新美齊澄清報載提及本公司停業及隱射屬大同集團上市櫃公司之報導。。1.報導內容：大同集團上市櫃公司，如果扣除停業的新美齊，總共有7家企業，除了大同、大同新境界外，幾乎都。變成殘值，小股東若把畢生積蓄用來投資大同的股票，很多已形同壁紙.....。(載自蘋果日報)。2.澄清說明：新美齊針對蘋果日報提及本公司停業及隱射屬大同集團上市櫃公司之報導，係屬媒體誤報，本公司嚴。正澄清本公司非屬大同集團之轉投資公司。財務業務整體營運獨立而且營運正常。。【TCRI 觀點_20190109】。1.TCRI認為公司屬於大同集團;主係兩家經營團隊(最終控制者)具2至3親等之關係.。A.新美齊(2442)雖於2018/12/19及2019/1/1兩度發重訊澄清-公司非大同(2371)之轉投資公司;但不是大同轉投資。不代表就不是大同集團.。B.新美齊董事長林傳捷,與董事林傳凱(林傳捷兄弟)的祖父是林挺生(大同創辦人林尚志之子),而大同董事長林郭。文艷配偶林蔚山的父親也是林挺生;另新美齊持股13.93%,及持股12.97%之前兩大股東高美投資,及高輔投資負責人。是林傳捷母親溫靜如,及溫惠中.。C.只不過林挺生兩房子孫內鬥時有所聞,這或許是新美齊不希望被市場認為是大同集團的原因.。

小組分工

劉品妤：AWS 與 Wordpress 網頁製作、爬蟲最新新聞資料

王昱達：股價資料整理、資料庫建立、網頁串接

楊廣元：新聞評分與股價預測模型的建立與調整參數

呂明諺：切詞與進一步的優化

Trello連結:

<https://trello.com/b/sVGChbne/%E7%8E%89%E5%B1%B1%E7%AC%AC%E4%B8%80%E9%A1%8C%EF%BC%8D%E5%8A%89%E5%93%81%E5%A6%A4>

感謝聆聽