

# Team 33



組員：

邱祥鴻：臺大財務金融所

黃柏森：東吳資料科學系

陳星儀：東吳資料科學系

洪鈺嫻：東吳企業管理學系

南山人壽題目一  
「文本生成--市場焦點機器人」

# Checkpoint3



# CONTENT 目錄

01 / 分工內容

02 / 專案故事流程

03 / 會議進度與業師反饋

04 / 預期輸出架構

05 / 階段性實驗成果

06 / 問題與困難

# 分工內容

邱祥鴻

新聞資料爬蟲、Google聲量  
趨勢串接、統整專案程式碼

黃柏森

知識圖譜模型建構、網頁架構

陳星儀

自動摘要模型建構、網頁架構

洪鈺姍

股價趨勢、CAPM模型運算、  
PPT輸出

PART

01

# 專案故事流程

PART

02

由於專案方向類似驗證關鍵字的個股是否值得投資，所以一整個專案中會進行三個階段的驗證。

1. Google聲量未來趨勢。(只取上升、持平)
2. 股價走勢趨勢。
3. CAPM模型的必要報酬率。

# 使用情境故事流程

第一驗證



第二驗證  
第三驗證

# 會議進度與反饋

## PART 03

10	📄 第七次實體會議	2021/05/13 4:00 PM-7:00 PM	整體架構	策劃 討論
11	📄 第四次線上會議	2021/05/20 3:00 PM-6:30 PM	整體架構	策劃 討論
12	📄 第五次線上會議	2021/05/27 3:00 PM-6:30 PM	整體架構 程式整合	策劃 討論

1

業師對於原專案流程的投資面資訊有疑慮：

1. 壽險機構投資策略並非是以短期利益為主，故對於波動度資訊需求低。
  2. 相關技術指標參考性較低，其投資是以長期利益或是未來潛力為主。
- ✓ 討論過後→提供一條專案流程的新思路，以大型機構投資角度看待使用需求，利用google trend找尋焦點新聞中一段期間內相對聲量持續上升的熱門主題，進而找出未來可以流行的主題趨勢。

2

利用news-api爬取新聞。進度狀況與難題：

1. 資料上限100筆，樣本過小，對於後續圖譜的模型難以訓練。
  2. 若採用自動輸入，新聞議題會過於廣泛，無法提取特定關鍵字。
- ✓ 討論過後→目前新增爬google news的新聞內容。

3

利用tf-idf的方法來生成摘要。進度狀況與難題：

1. 若整合大量單篇新聞，前後文語意不通順，難以成章。
- ✓ 討論過後→嘗試使用bart的方法來生成，但如果時間緊迫，就以產出最有關聯性的摘要為優先。

4

關聯性的模型建構。進度狀況與難題：

1. 尚未透過NER識別出一些人名、地名等類似的自定義類別。
2. 目前的模型尚無法去除文本中的很多無意義詞語。
3. 尚未達到判斷企業名稱後自動連結企業的Ticker。

□ 待討論解決。

5

利用股價計算CAPM。進度狀況與難題：

1. 利用OLS模型計算Beta值，判定係數普遍不高。
  2. 正進行完整運算：加入期望值、無風險利率、個股報酬等參數。
- ✓ 討論過後→增加使用Fama-French三因子模型測試。



# 預期專案架構

## PART 04

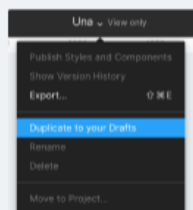
# 預期簡報輸出架構

## Presentation Template

by templatery.co

### HOW TO CUSTOMIZE THE PRESENTATION

You need to have a Figma account (it's free).  
Then simply export this file or duplicate it.  
You can use the dropdown in the middle of the screen.



### HOW TO PRESENT

1. Click on **Prototype** in top right
2. Wire your slides together ([tutorial](#))
3. Click the **Play** button

### ABOUT TEMPLATERY.CO

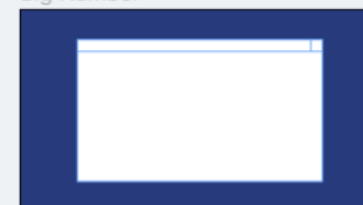
If you want to take a look at other presentation crafted with love ❤️ head to

[Website](#)  
[Twitter](#)

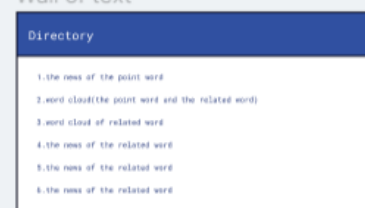
#### Cover



#### Big Number



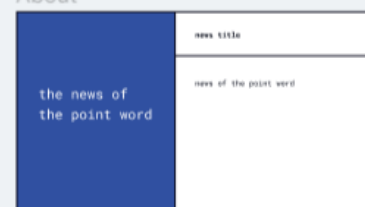
#### Wall of text



#### Three points



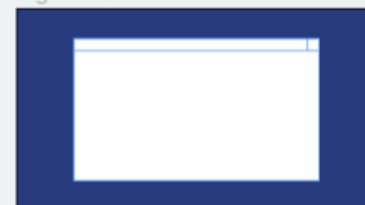
#### About



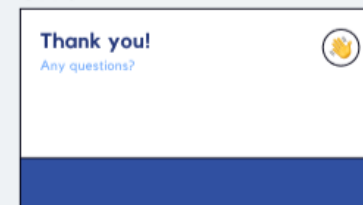
#### Three points



#### Big Number



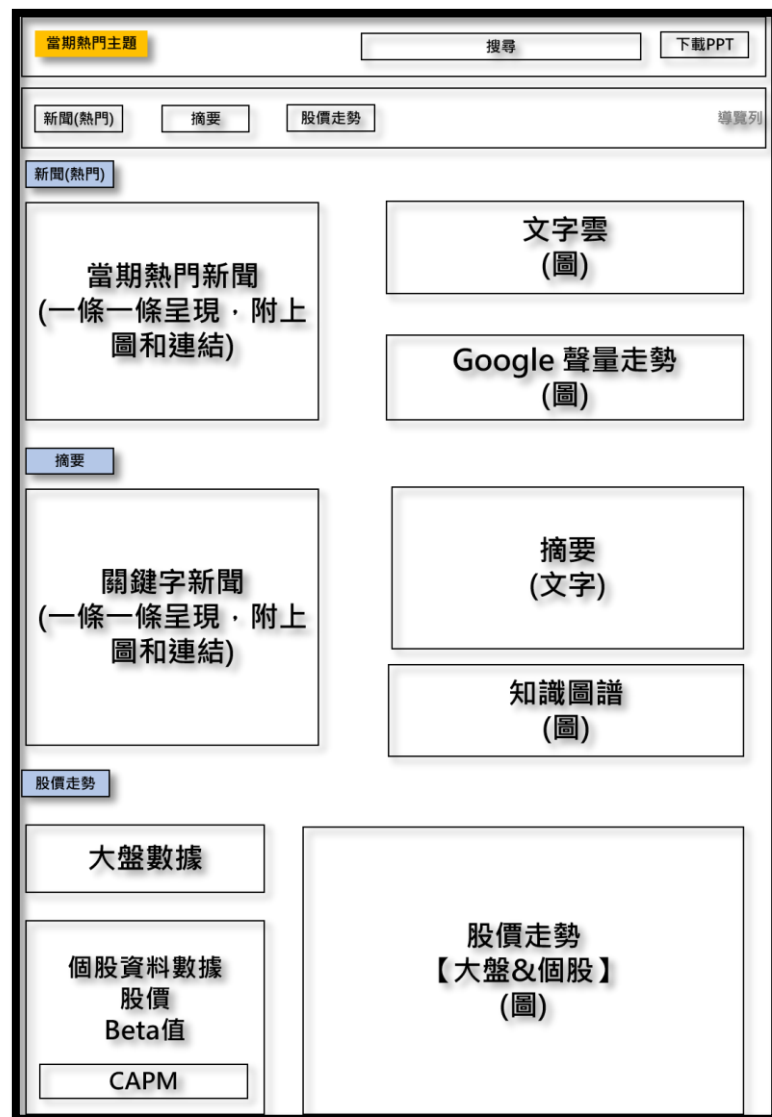
#### Outro



【正在動工...】

基本上以商務簡約風為主體架構，內容涵蓋熱門文字雲圖、Google聲量圖、關鍵字及關聯摘要、股價趨勢圖等本次專案生成的所有資訊。

# 預期簡報輸出架構



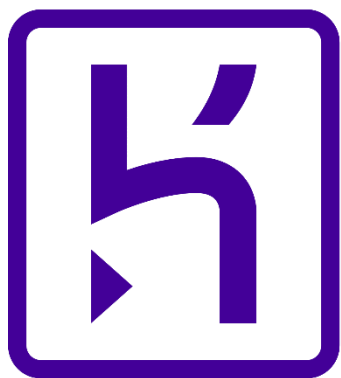
基本上以單頁可滑動的網頁為主。網頁預設的內容是以當期熱門新聞全自動產生的相關內容為展示，並於最上層設有可彈性搜尋的關鍵詞，以利即時搜尋使用，且設有一鍵下載的簡報生成功能。

可能會調整新聞的版面，讓使用者可以先預覽新聞，若是要觀看全文，利用連結跳轉自新聞原聞處。一旦點開新聞，文字雲圖和google聲量趨勢圖會向下排版。

中間可以選擇想查看的主軸，目前分為新聞、摘要及股價走勢。

(後續若有新增的圖文或功能，可能會些許微調)

# 預期雲端部署架構



# HEROKU

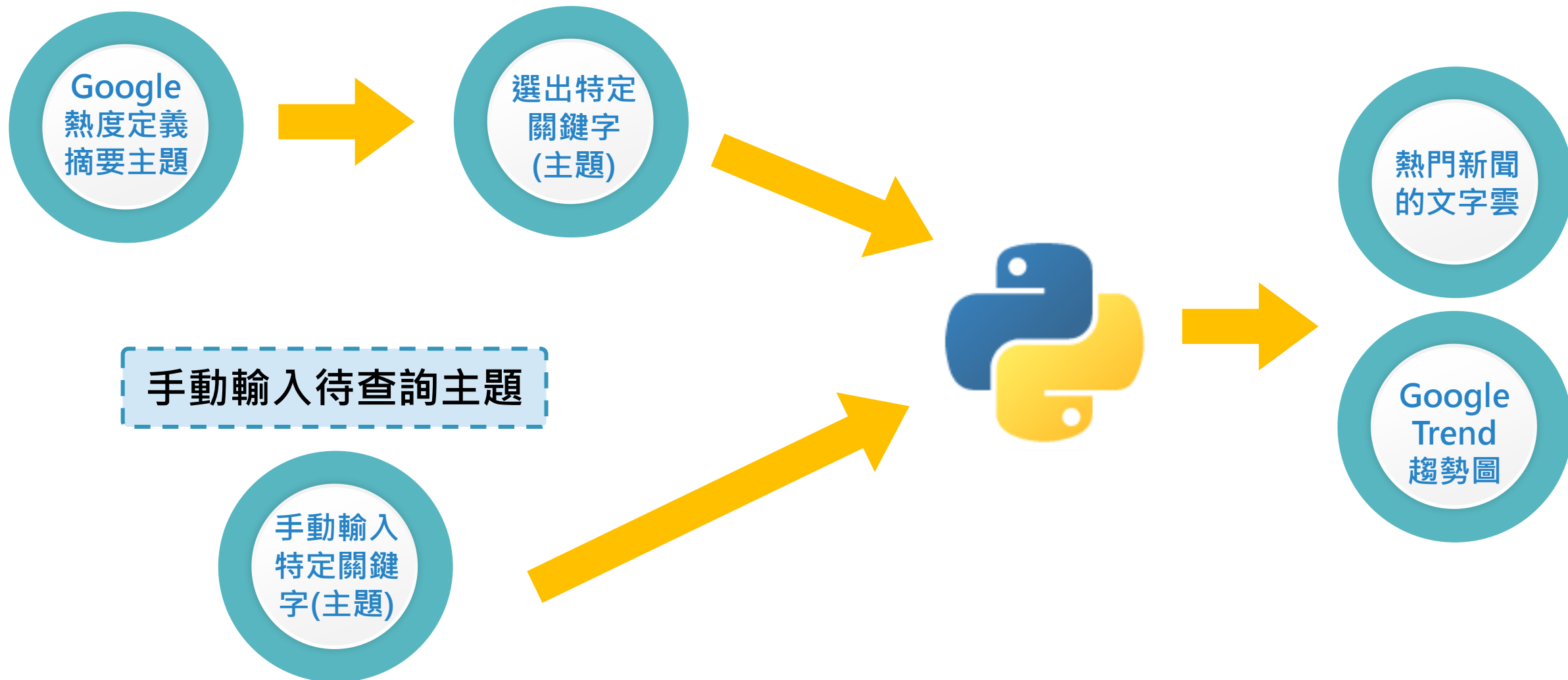
目前討論較偏向採用Heroku app的方式進行雲端部署，相關的專案整合架構主要是依照前述專案流程的方式進行整合。

由於各組員都還在進行最後的功能調整，故預計於下周起(5/31)由邱祥鴻同學進行專案的功能程式的整合以及網頁的製作，並暫訂下周 (6/11)進行雲端部署的測試。

# 階段性實驗成果

## PART 05

自動產生趨勢主題



特定字詞：Bitcoin

# news-api 特定字詞 新聞內容爬取 (Test-Data)

```
1 {
2   "status": "ok",
3   "totalResults": 5510,
4   "articles": [
5     {
6       "source": {
7         "id": "reuters",
8         "name": "Reuters"
9       },
10      "author": "Oliver Griffin",
11      "title": "UPDATE 1-Colombia will work to keep investor
12              confidence after ratings cut -minister - Reuters",
13      "description": "Colombia will work to maintain market
14                    confidence in its finances, Finance Minister Jose Manuel
15                    Restrepo said on Wednesday, after S&P Global Rating
16                    relegated its credit rating to junk bond status.",
17      "url": "https://www.reuters.com/article/colombia-economy
18             -idUSL2N2N703M",
19      "urlToImage": "https://s1.reutersmedia.net/resources_v2
20                    /images/rcom-default.png?w=800",
21      "publishedAt": "2021-05-20T02:33:00Z",
22      "content": "(Adds context on credit downgrade)\r\nBOGOTA,
23                 May 19 (Reuters) - Colombia will work to maintain market
24                 confidence in its finances, Finance Minister Jose Manuel
25                 Restrepo said on Wednesday, after S&P... [+2114 chars]"
26    },
27    {
28      "source": {
```

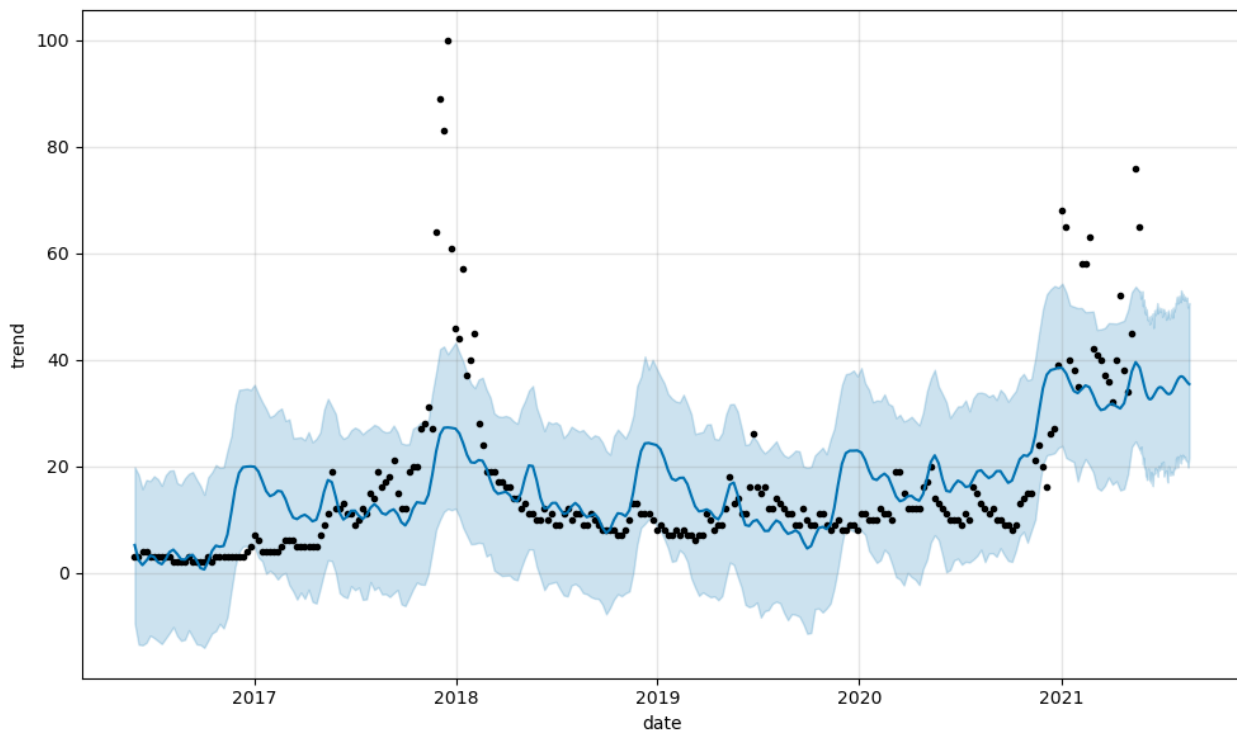
# WordCloud

## 特定字詞輸出

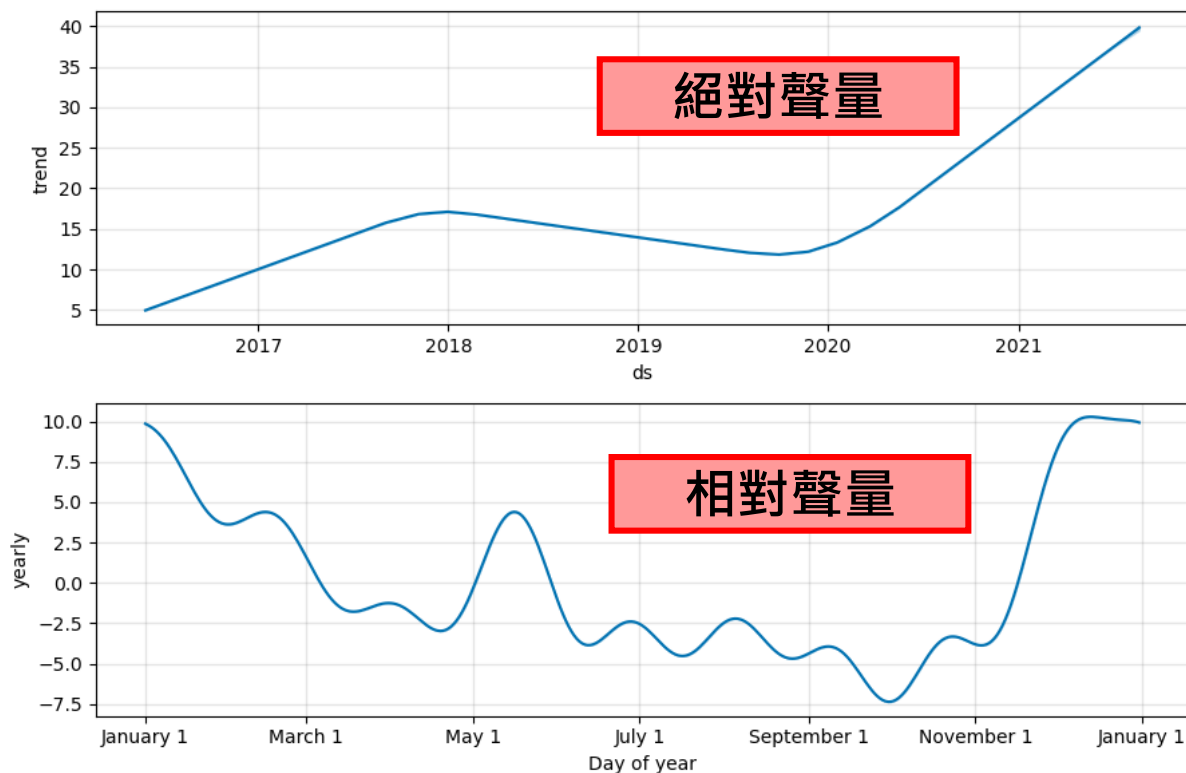


特定字詞：Bitcoin

## Google Trend特定字詞 之 聲量信心區間及趨勢圖



## Google Trend特定字詞 之 絕對聲量及相對聲量趨勢圖





建立半人工關聯式知識圖譜

串接前述  
的Test-  
Data

透過大量文本尋找關聯詞

輸入關鍵  
字(主題)



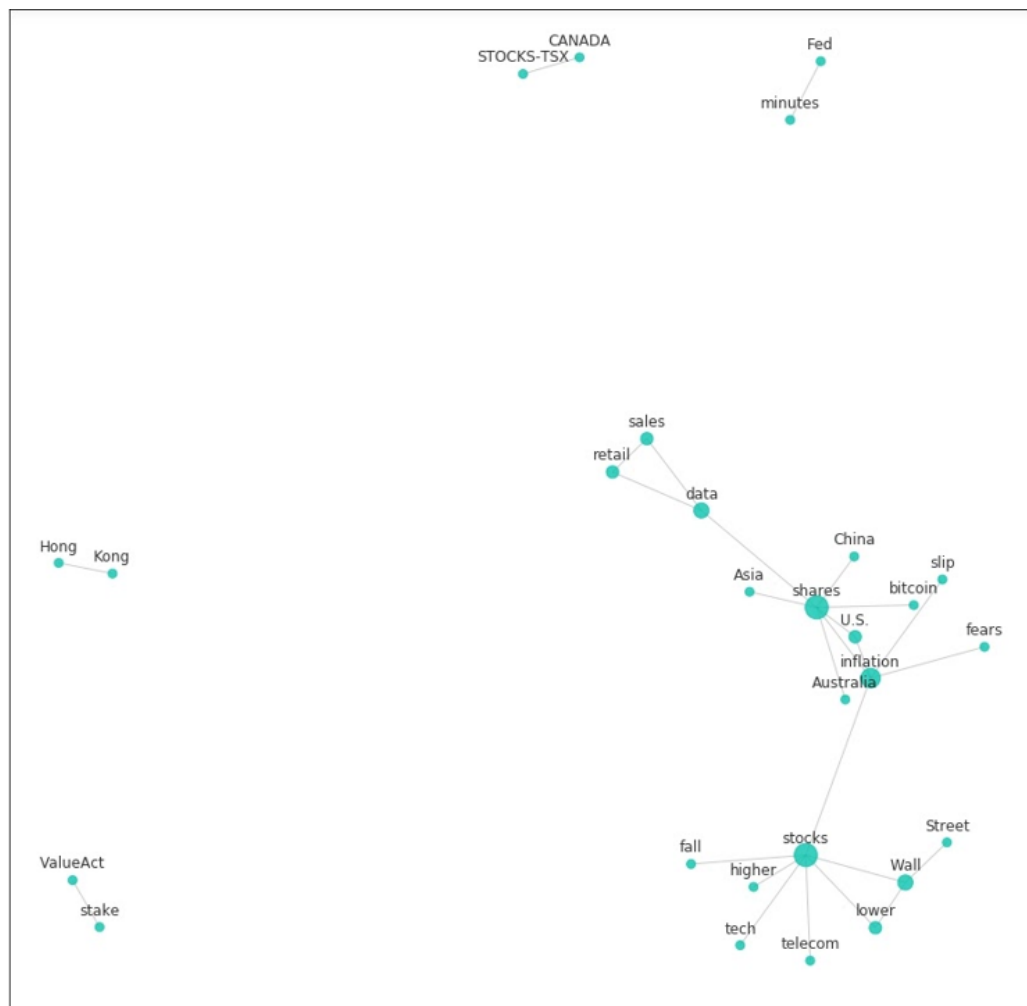
共現性資  
料視覺化

輸出相關  
關聯詞

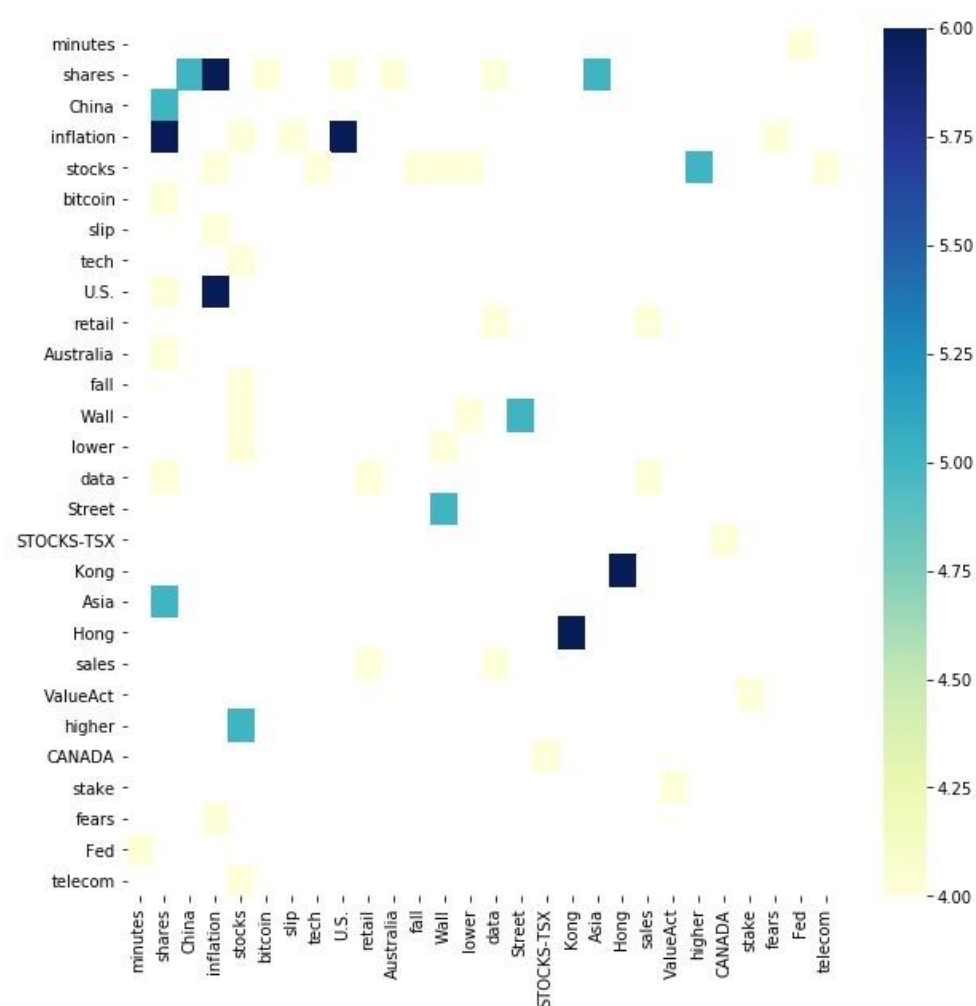
特定字詞  
Bitcoin

將蒐集到的新聞資料整理成共現矩陣，再透過調整出現頻率進而調整矩陣的大小，最後將結果以視覺化方式呈現。

## 特定字詞的共現圖



## 特定字詞的熱點圖



新聞的主題字詞：Bitcoin

查找的特定字詞：inflation

## 特定字詞的關聯詞

Please enter keyword : inflation

[['shares', 6], ['U.S.', 6], ['slip', 4], ['fears', 4], ['stocks', 4]]

[關聯詞, 出現次數]

將前述的Test-Data丟入模型中，並假設查找的特定字詞為inflation。  
進而產生的關聯詞有shares、U.S.、slip、fears、stocks。

取得特定關鍵詞的新聞內容後

取得特定  
關鍵字  
(主題)

取得相關關聯詞的新聞內容後

取得相關  
關聯詞



關鍵字的  
主題摘要

多個關聯  
詞的各篇  
摘要

## 特定字詞（文本主題）：Bitcoin

### 自動摘要

原文

Colombia will work to maintain market confidence in its finances, Finance Minister Jose Manuel Restrepo said on Wednesday, after S&P Global Rating relegated its credit rating to junk bond status. Global ratings agency S&P on Friday affirmed its long-term sovereign credit rating on Israel at "AA-/A-1+" and maintained its outlook at "stable". Investment firm ValueAct liquidated its stake in Morgan Stanley (MS.N) and trimmed holdings in Citigroup Inc (C.N), KKR & Co Inc (KKR.N) and SLM Corp (SLM.O) in early 2021, pointing to a shift away from some financial firms that it has owned for several years... South Africa's rand ticked higher on Thursday ahead of a central bank meeting, while most other emerging market currencies rose slightly, even as hints of eventual policy tightening from the Federal Reserve clouded their outlook. With the possibility of an inflation-led market sell-off on their minds, many bond investors are retreating for shelter to the unlikely haven of junk bonds, a sector typically considered one of the riskiest. With the possibility of an inflation-led market sell-off on their minds, many bond investors are retreating for shelter to the unlikely haven of junk bonds, a sector typically considered one of the riskiest. China's yuan traded largely flat against the dollar on Monday, as support from a firmer official currency guidance rate and tighter cash conditions in interbank markets made up for weakness from disappointing

從原文中抽取摘要的程度

0.20

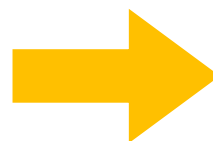
0.10

1.00

## 文本主題摘要生成

FUNDAMENTALS Spot gold was up 0.1% at \$1,844.60 per ounce b... Asian share markets turned mixed on Monday as data on Chinese retail sales missed expectations though industrial output stayed solid, while more evidence of global inflation pressures helped gold to a three-month peak. A number of well-known U.S. hedge funds bought value stocks and blank-check acquisition companies, selling some winners from the technology-led stock rally as bond yields rose during the first quarter, filings released on Monday showed. A number of well-known U.S. hedge funds bought value stocks and blank-check acquisition companies, selling some winners from the technology-

取得主題企業  
抑或式關聯詞企業的ticker



個股ticker : AAPL  
( 個股 : Apple Inc. )

## 單一個股的股價趨勢圖

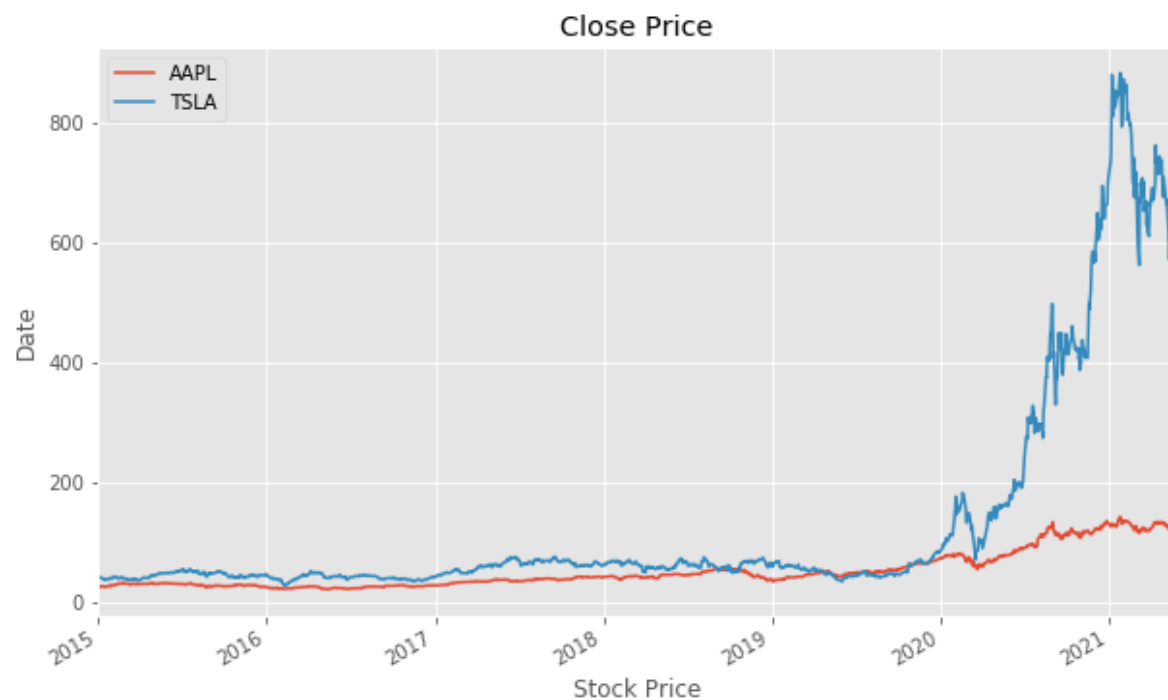


## 單一個股的企業基本資料

```
{ 'zip': '95014',  
  'sector': 'Technology',  
  'fullTimeEmployees': 100000,  
  'longBusinessSummary': 'Apple Inc. designs, manufactures, and markets various related services. The company offers iPhone, a line of wearables, home, and accessories comprising AirPods, Apple TV, Apple Music, and more. It also provides AppleCare support services; cloud services such as iCloud to discover and download applications and digital content, such as books, movies, and TV shows; and other services, such as Apple Arcade, a game subscription service; Apple Music, which provides a subscription news and magazine service; Apple TV+, which offers exclusive content, as well as licenses its intellectual property. The company serves government markets. It sells and delivers third-party applications for its products through the App Store and online stores, and direct sales force; and third-party cellular service providers. Apple is headquartered in Cupertino, California.',  
  'city': 'Cupertino',  
  'phone': '408-996-1010',  
  'state': 'CA',  
  'country': 'United States',  
  'companyOfficers': [],  
  'website': 'http://www.apple.com',  
  'maxAge': 1,  
  'address1': 'One Apple Park Way',  
  'industry': 'Consumer Electronics',  
  'previousClose': 126.9,  
  'regularMarketOpen': 126.955,
```

個股ticker：AAPL、TSLA、^GSPC  
( 個股：Apple Inc.、Tesla、S&P500 Index )

## 多檔個股的股價趨勢圖



## 多檔個股&大盤的股價趨勢圖





個股ticker : AAPL、^GSPC、^IXIC  
( 個股 : Apple Inc.、S&P500 Index、NASDAQ Index )

OLS :

普通最小平方線性迴歸

OLS  
(Apple fit S&P500 Index)

判定係數=5成  
( 擬合程度、解釋力 )

OLS  
(Apple fit NASDAQ Index)

OLS Regression Results

Dep. Variable: Apple

Model: OLS

Method: Least Squares

Date: Sat, 29 May 2021

Time: 04:31:54

No. Observations: 78

Df Residuals: 76

Df Model: 1

Covariance Type: nonrobust

R-squared: 0.458

Adj. R-squared: 0.451

F-statistic: 64.27

Prob (F-statistic): 1.02e-11

Log-Likelihood: 112.34

AIC: -220.7

BIC: -216.0

α值皆>0

⇒顯示有超額報酬

	coef	std err	t	P> t	[0.025	0.975]
const	0.0092	0.007	1.376	0.173	-0.004	0.022
S&P500 Index	1.1742	0.146	8.017	0.000	0.883	1.466

Omnibus: 1.307

Prob(Omnibus): 0.520

Skew: -0.090

Kurtosis: 3.432

Durbin-Watson: 2.095

Jarque-Bera (JB): 0.711

Prob(JB): 0.701

Cond. No.: 22.3

$\alpha$ 值皆>0  
⇒顯示有超額報酬

OLS Regression Results

Dep. Variable:

Apple

Model:

OLS

Method:

Least Squares

Date:

Sat, 29 May 2021

Time:

04:12:26

No. Observations:

78

Df Residuals:

76

Df Model:

1

Covariance Type:

nonrobust

R-squared:

0.491

Adj. R-squared:

0.484

F-statistic:

73.21

Prob (F-statistic):

9.46e-13

Log-Likelihood:

114.75

AIC:

-225.5

BIC:

-220.8

	coef	std err	t	P> t	[0.025	0.975]
const	0.0056	0.007	0.864	0.390	-0.007	0.01
NASDAQ Index	1.0648	0.124	8.556	0.000	0.817	1.31

Omnibus:

2.055

Durbin-Watson:

1.978

Prob(Omnibus):

0.358

Jarque-Bera (JB):

1.374

Skew:

0.216

Prob(JB):

0.503

Kurtosis:

3.485

Cond. No.

19.5

Beta值皆相較大盤波動度較大

p-value皆<0.05⇒Beta係數顯著不為0

# 問題與困難

## PART 06

# 當前的困難（人、事、物）

## 新聞量不足

原先使用的new-api爬蟲所取得的新聞資料較少，對於知識圖譜的模型訓練較不足，故要多新增爬取google新聞以確保足量的文本來源。

## 摘要模型精確度低

整合大量單篇新聞後，發現模型產出的摘要前後文語意不通順，難以成章。離預期的摘要文本還有距離，目前正更換方法，希望可以提高摘要模型的精確度。



## 關聯性模型些微雜訊

目前尚未透過NER識別出自定義類別的詞語，因此如前述的文本中的Hong Kong被分成兩個字，尚且未能完全去除許多無意義的詞語。

## 半數組員狀態不好

組內人本來就少，半數組員的身心狀況近期都不大好，故此進度稍嫌落後，已載加緊趕工。

# Team 33



**謝謝聆聽**