

Knowledge graph building & Information extraction

個股新聞報導之間的相似度

台大會計碩二 劉智玄
台大經濟碩一 高翊傑
台大財金四 楊凱鈞
東吳巨資三 周柏佳

指導業師：中信銀行-林志豪
指導老師：張智星 老師
蔡芸琿 老師

問題描述

- 欲檢驗各大新聞網站中，個股新聞報導用詞的相似度，是否能呈現在股價中
- 預期個股之間新聞詞組相似度越高，股價走勢也會有一定的相似度

目標是以消息面出發建立類股知識圖譜

資料集樣態說明

- **四家網路新聞：**
ETtoday新聞雲、自由時報電子報、中時電子報、東森新聞
- **新聞選取：**
一個月內，交易金額前20大之半導體類股相關新聞

解決方式之流程

- 從各家新聞網站中爬取一個月內目標個股的相關新聞，並使用結巴套件做斷詞以及詞組清理
- 將詞組利用word2vector套件訓練後轉為詞向量
- 將得到的高維度詞向量利用TSNE降維至二維平面並算出詞向量之間的相似度矩陣
- 檢視此相似度矩陣和實際股價間的差異



DEMO & RESULTS

相似度二維分布



20檔股票之新聞詞向量二維分布圖：

- 距離越近之個股其詞向量間有較高之相關性
- 台積電與其他個股的詞向量相關性均不高
- 新聞量較大之個股(例如：台積電、聯發科、聯電)有較多雜訊，因此會距離其他個股較遠

Telegram Chat Bot



按下START啟動聊天機器人後
會自動傳送可選取的個股

Telegram Chat Bot

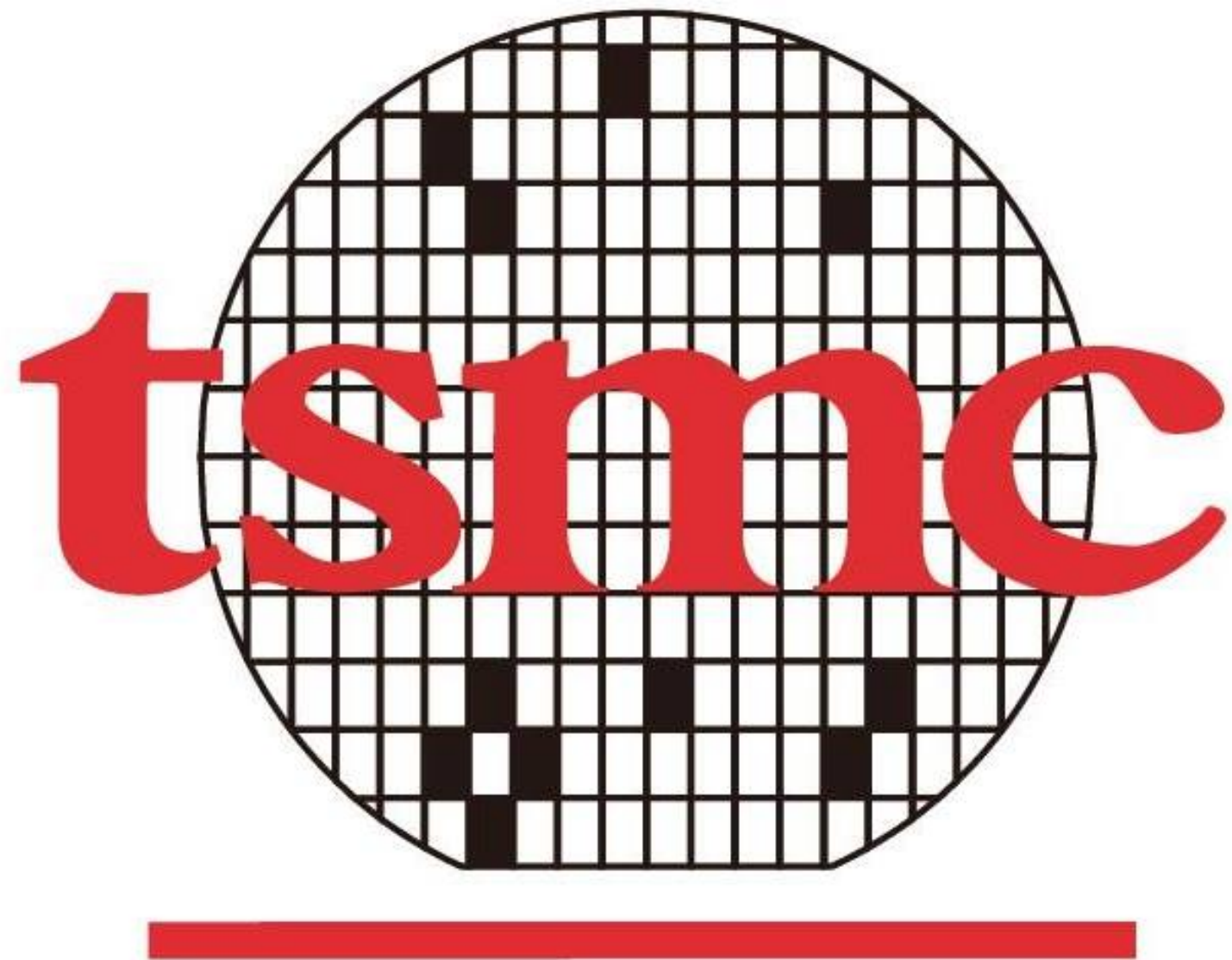


直接點選對話框中個股連結即可，另外設有防呆機制

按下個股連結後
聊天機器人會自動傳送個股之相似度圖譜
圖譜中會顯示下列資訊：

- 相似度前五檔個股
- 相關係數
- 個股之前日開盤、收盤、最高、最低價

挑選解說標的



主要新聞

華為禁令

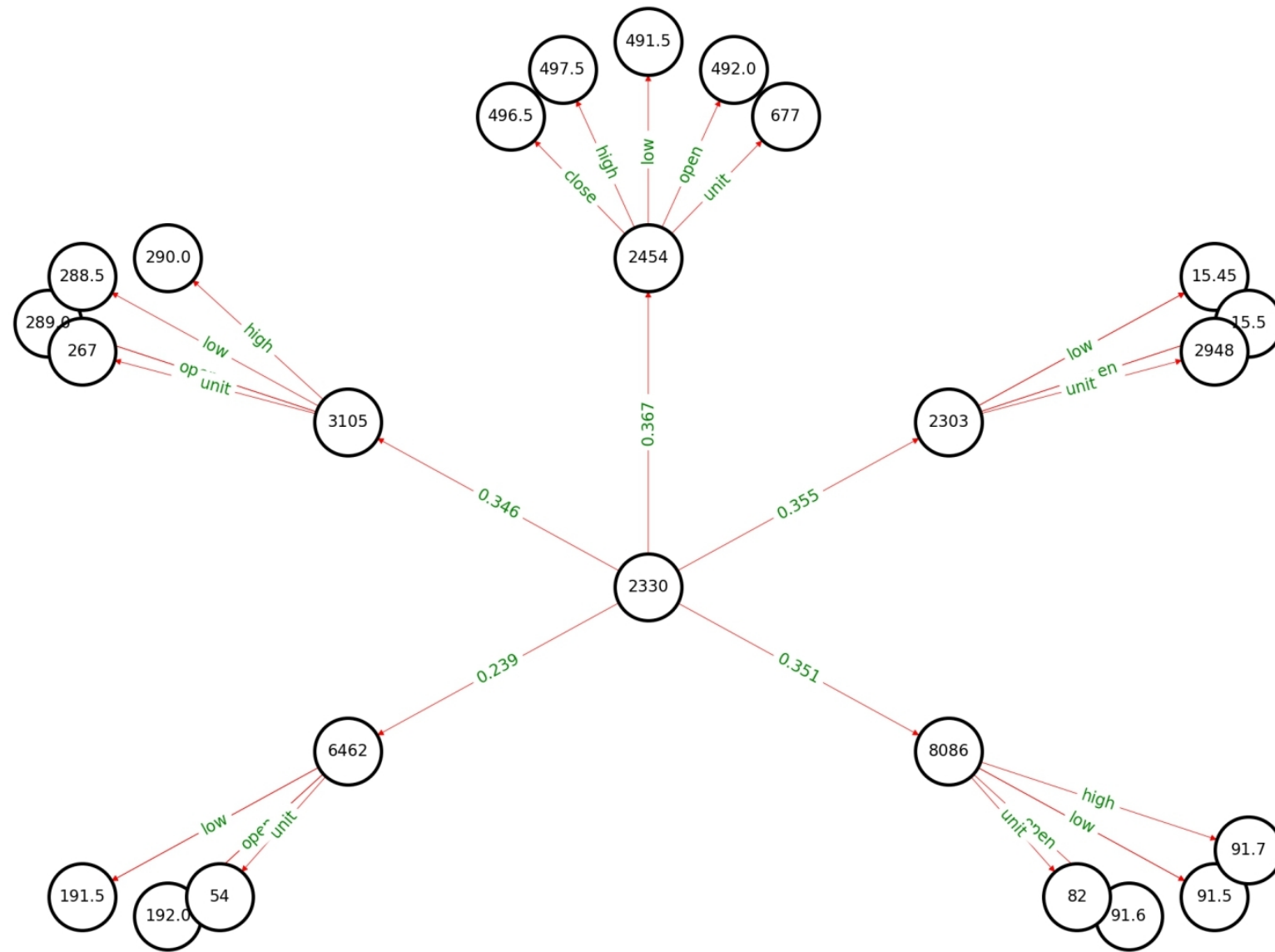
特性

相關係數普遍偏低且同時有正負相關

主要業務

製造與銷售積體電路以及其他晶圓半導體裝置。提供前述產品之封裝與測試服務、積體電路之電腦輔助設計技術服務。提供製造光罩及其設計服務。

台積電 - 相似度前五檔分析



主要新聞

3105：穩懋

/ 華為禁令

2454：聯發科

/ 華為禁令

2303：聯電

/ 華為禁令

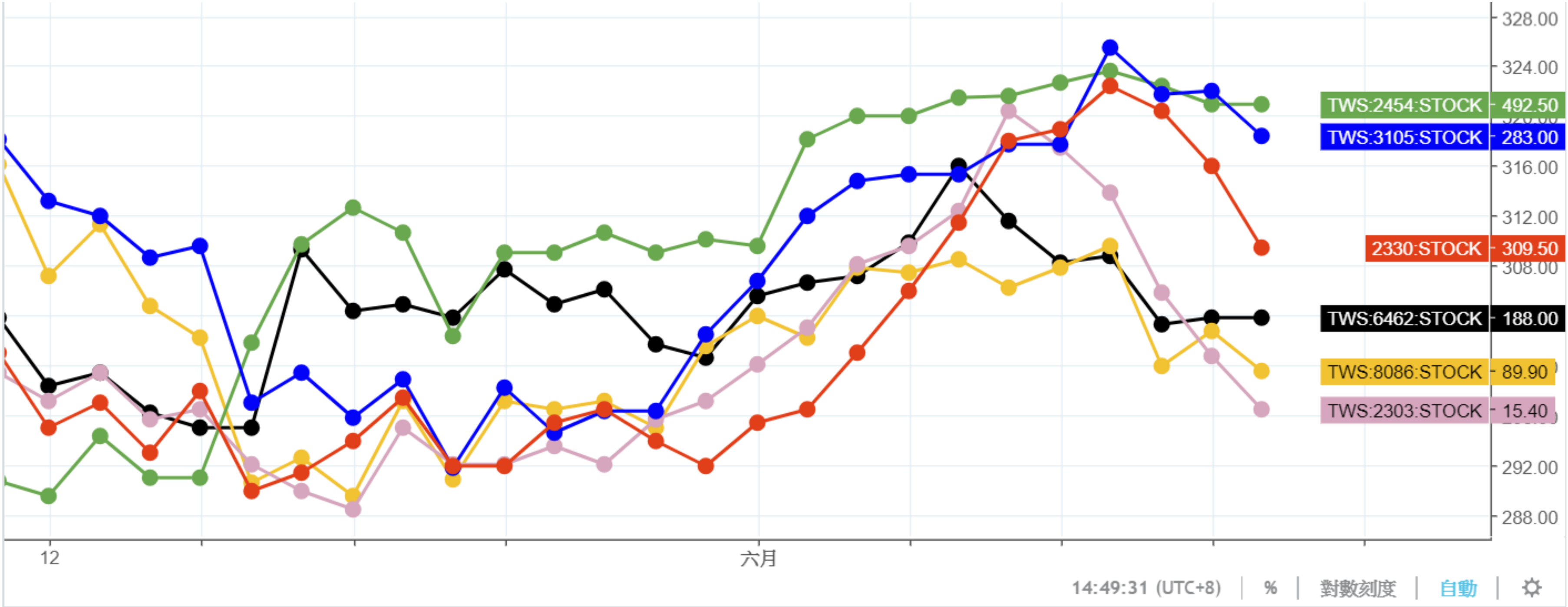
8086：宏捷科

/ 華為禁令

6462：神盾

/ 華為禁令

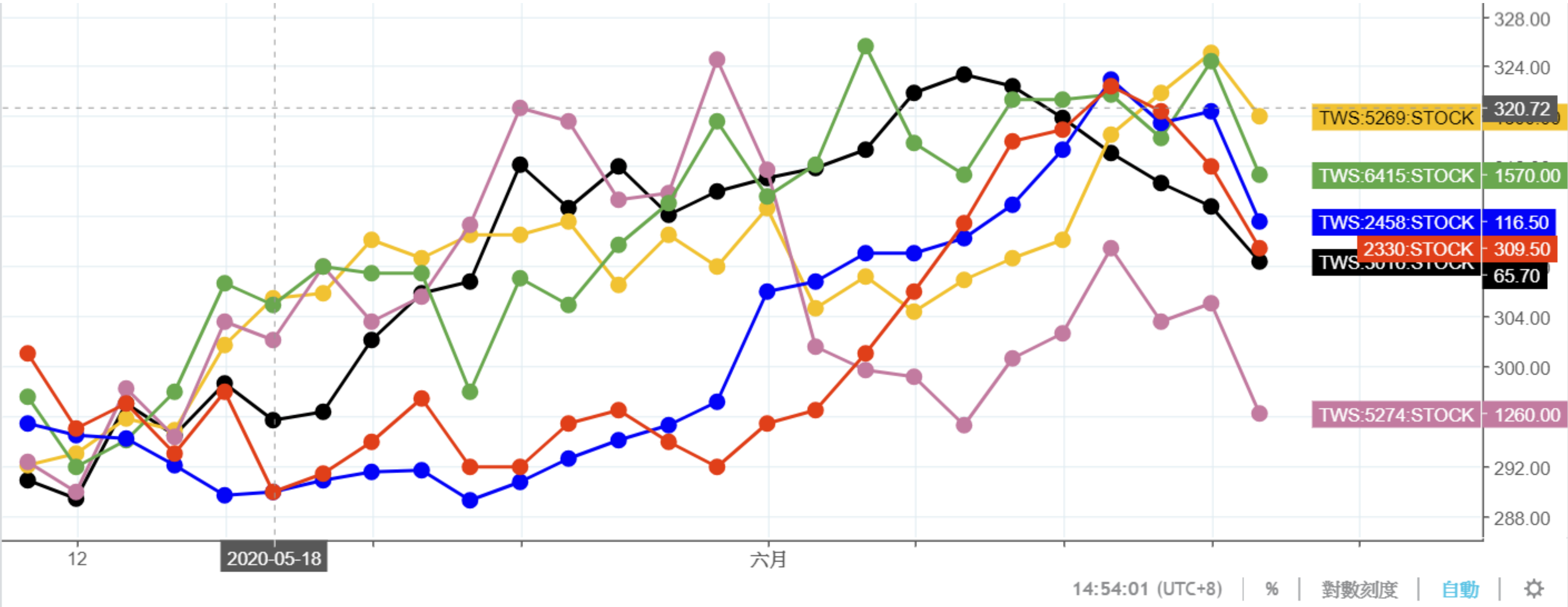
台積電 - 相似度前五檔股價走勢



台積電 - 相似度前五檔股價相關係數矩陣

	2330	3105	2454	2303	8086	6462
2330	1.000					
3105	0.595	1.000				
2454	0.078	0.114	1.000			
2303	0.656	0.465	0.027	1.000		
8086	0.562	0.801	0.108	0.536	1.000	
6462	0.303	0.404	0.444	0.116	0.477	1.000

台積電 - 相似度後五檔股價走勢



台積電 - 相似度後五檔股價相關係數矩陣

	2330	2458	6415	5274	5269	3016
2330	1.000					
2458	0.468	1.000				
6415	0.336	0.191	1.000			
5274	0.203	0.084	0.443	1.000		
5269	0.193	0.263	0.115	0.428	1.000	
3016	0.450	0.081	0.339	0.385	0.005	1.000

主要貢獻與改進方向

主要貢獻

透過直覺的聊天機器人幫助投資人挑選具有一定相似程度的相關個股：

- 透過個股新聞詞組分析，能夠篩選出前五檔股價走勢相關個股，提供受同一事件影響之其他個股，方便使用者進行進一步之投資策略研擬。
- 提供篩選出之前五檔個股之股價資訊，方便使用者初步判斷股價水位。
- Telegram的聊天機器人提供友善使用者的應用介面，並內建個股連結，可快速取得投資所需資訊。

改進方向

- 目前僅以近一個月交易金額前20大之半導體類股新聞做實測，預計未來納入完整資料庫後即可上線使用。
- 目前係以單篇新聞為單位做斷詞，未來會以新聞中之單段描述為單位進行斷詞，預計可以達到更好的相似度成果。
- 遠程目標為建立台灣所有上市櫃公司相似度之知識圖譜，即可以一張圖快速了解所有公司之消息面相關性。

分工任務

高翊傑：爬蟲、word2vector建模、TSNE視覺化、similarity矩陣建置
周柏佳：主題發想、爬蟲、資料清整、圖譜實作、聊天機器人實作
楊凱鈞：主題發想、爬蟲、財金相關建議、投影片製作、介紹影片錄製
劉智玄：主題發想、爬蟲、成果分析、投影片製作、介紹影片錄製

作品連結

專題作品的測試連結與說明：

https://t.me/Fintech_graph_bot

專題 GitHub：<https://github.com/rebeca0521/CTBC-Project>

影片超連結：

THANK YOU!