



致理科技大學

資訊管理系專題報告

馬斯克推文之情感分析 Elon Musk Twitter Sentiment Analysis

專題生：
(10710147)張雅喬
(10710149)林以晴
(10710108)陳婕名
(10710115)黃乙茜
(10710120)施鈺函
(10710126)謝涓瑾

指導教授：王嫻惠 老師

中華民國 111 年 05 月

致理科技大學

資訊管理系

畢業專題

馬斯克推文之情感分析

一一〇學年度

致理科技大學 專題報告審核書

本校 資訊管理 系（所）林以晴(10710149)、
張雅喬(10710147)、黃乙茜(10710115)、施鈺
函(10710120)、謝滄瑾(10710126)、陳婕名(10
710108)等君所提論文馬斯克推文之情感分析

Elon Musk Twitter Sentiment Analysis

經本委員會審定通過，特此證明。

口試委員會

委員： _____

指導教授： _____

系主任： _____

中華民國 111 年 05 月

致理科技大學

授權書

本授權書所授權之專題報告在致理科技大學

110 學年度第一學期所撰寫。

專題名稱：馬斯克推文之情感分析-Elon Musk Twitter Sentiment Analysis

本人具有著作財產權之論文或專題提要，授予致理科技大學，得重製成電子資料檔後收錄於該單位之網路，並與台灣學術網路及科技網路連線，得不限地域時間與次數以光碟或紙本重製發行。

本人具有著作財產權之論文或專題全文資料，授予教育部指定送繳之圖書館及本人畢業學校圖書館，為學術研究之目的以各種方法重製，或為上述目的再授權他人以各種方法重製，不限時間與地域，惟每人以一份為限。並可為該圖書館館藏之一。

本論文或專題因涉及專利等智慧財產權之申請，請將本論文或專題全文延至民國 111 年 7 月 1 日後再公開。

上述授權內容均無須訂立讓與及授權契約書。依本授權之發行權為非專屬性發行權利。依本授權所為之收錄、重製、發行及學術研發利用均為無償。

(上述同意與不同意之欄位若未勾選,本人同意視同授權)

同意 不同意

學生簽名：

(親筆正楷簽名)

指導老師姓名：

(親筆正楷簽名)

中華民國 111 年 05 月

摘要

專題報告名稱：馬斯克推文之情感分析-Elon Musk Twitter Sentiment Analysis

頁數：46

校系別：致理科技大學資訊管理系

完成時間：110 學年度第 2 學期

專題生：林以晴、張雅喬、黃乙茜、施鈺函、謝涓瑾、陳婕名

指導教授：王嫻惠 副教授

關鍵詞：情感分析、自然語言、機器學習、深度學習

隨著科技及網路的進步，網路貨幣已是現在的投資趨勢。加上世界疫情肆虐，許多人不得不保持社交距離並待在家中落實防疫，待在家工作的同時也代表網路逐漸成為人們不可或缺的一部份。如今社群軟體的發達，導致許多網路用戶會在社群上發表個人言論，有些言論甚至會間接影響到國情、政治、經濟的走向。尤其是名人在社群上發表的個人言論會對群眾的情緒產生影響，也間接造成了股市的漲跌。因此，本研究探討特斯拉之創辦人 Elon Musk 的 Twitter 推文言論是否會影響特斯拉股票漲跌。

ABSTRACT

Thesis Title : Elon Musk Twitter Sentiment Analysis

Pages : 46

University : Chihlee University of Technology

Graduate School : Department of Information Management

Date : May, 2022

Degree : Bachelor

Researcher : Chen, Jie-Ming, Chang, Ya-Chiao, Lin, Yi-Ching, Shih, Yu-Han, Hsieh, Yu-Chin, Huang, Yi-Qian

Advisor : Wang, Yen-Hui

Keywords : Sentiment Analysis, Natural Language, Machine Learning, Deep Learning

With the advancement of technology and the Internet, online currency has become the current investment trend. In addition to the raging epidemic all over the world, many people have to maintain social distancing and stay at home to implement epidemic prevention. Staying and working at home also means that the Internet has gradually become an indispensable part of people's life. Nowadays, the popularity of social software has caused many Internet users to make personal comments on the social network, and some comments even indirectly affect the national conditions, politics, and economic trends. In particular, the personal remarks made by celebrities on the social network will affect the mood of the masses and indirectly cause the stock market to rise or fall. Therefore, this study explores whether Tesla's founder Elon Musk's Twitter tweets will affect the rise and fall of Tesla's stock.

誌謝

首先，誠摯的感謝專題指導老師 王嫵惠 副教授以及班導師 曹祥雲 副教授。在兩位教授細心的指導和適時的給予建議下，使本組可以按步就班地完成作品。每當完成一個部分章節，會每週線上開會，老師會依目前完成的進度，給予一些待改進的建議，讓我們的系統更為完善。

在專題製作初期，組員們對於 Jupyter 這個環境是完全沒有接觸過的。在不熟悉的狀態下，也要感謝每位組員盡自己所能，利用課餘時間討論，共同學習。同學們也透過良好的溝通以及協作精神完成本專題。組員們更在這次專題製作中累積實作經驗，相信在未来應能夠應用在課業或職場上。



林以晴、張雅喬、黃乙茜、施鈺函、謝滄瑾、陳婕名 謹致
致理科技大學 資訊管理 學士班
中華民國 111 年 5 月

目錄

摘要	i
ABSTRACT	ii
誌謝	iii
目錄	iv
圖目錄	v
表目錄	vi
第壹章 緒論	1
第一節 研究背景	1
第二節 研究動機	1
第三節 研究目的	2
第貳章 文獻回顧與探討	3
第一節 深度學習 (Deep Learning)	3
第二節 長短期記憶模型 (LSTM)	5
第三節 學習預測標的	6
第參章 研究內容與方法	8
第一節 馬斯克個人介紹	8
第二節 研究架構	8
第三節 程式使用工具	9
第四節 研究內容	10
第五節 特斯拉新聞報導	10
第肆章 研究結果與設計	12
第一節 蒐集資料	12
第二節 LSTM-sentiment-analysis	13
第三節 運用深度學習訓練模型	16
第四節 研究結果	18
第伍章 結論與未來展望	23
第一節 結論	23
第二節 未來展望	23
參考文獻	25
附錄一 畢業專題 系統操作手冊	30

圖目錄

圖 1-1 川普針對伊朗之言論	1
圖 2-1 循環神經網路示意圖	3
圖 2-2 卷積神經網路示意圖	4
圖 2-3 LSTM 架構圖 1	5
圖 2-4 LSTM 架構圖 2	6
圖 2-5 計算方程式	6
圖 3-1 馬斯克個人照片	8
圖 3-2 研究架構圖	8
圖 3-3 Jupyter logo 圖	9
圖 3-4 TWINT logo 圖	9
圖 3-5 股價數據 1	10
圖 4-1 程式碼 1	12
圖 4-2 程式碼 2	12
圖 4-3 股價數據 1	13
圖 4-4 整理數據 1	13
圖 4-5 程式碼 3	14
圖 4-6 訓練執行結果 1	14
圖 4-7 訓練執行結果 2	15
圖 4-8 訓練執行結果 3	15
圖 4-9 程式碼 4	15
圖 4-10 程式碼 5	16
圖 4-11 程式碼 6	17
圖 4-12 訓練執行結果 3	17
圖 4-13 程式碼 7	18
圖 4-14 程式碼 8	18
圖 4-15 程式碼 9	19
圖 4-16 程式碼 10	19
圖 4-17 程式碼 11	19
圖 4-18 程式碼 12	20
圖 4-19 驗證結果 1	20
圖 4-20 驗證結果 2	21
圖 4-21 預測結果 1	22

表目錄

表 1-1. 主要股價指數、債券、原物料與匯率表現	2
表 2-1. 長句子拆成有意義的單詞	4



第壹章 緒論

第一節 研究背景

當今社會社群軟體越來越發達，不論 Facebook、Instagram 以及 Twitter，都是近年來熱門的社群軟體。越來越多網路用戶在社群軟體上發表個人言論。事件的發展往往會因為各種不同的言論，而間接影響到國情、政治，甚至於經濟走向。因此現今越來越多人重視網路輿論，收集網路正反言論，使大數據已成為目前的國際趨勢走向。

舉特斯拉汽車執行長馬斯克的網路言論為例。馬斯克在今年 2 月份花費了 15 億美元以企業名義購入比特幣，接著又在 3 月宣布開放比特幣購車，但馬斯克卻於 5 月時在 Twitter 發佈一篇貼文表示，不管是比特幣不管是挖礦還是交易過程都有助化石燃料消耗造成過多碳排放量，因此取消了比特幣購車。根據 CoinDesk 報價顯示，發文隔天收盤價大跌 8.5%，比特幣價格降到 34,877 美元。另外，從數位時代(2020)的網站能得知，6 月 13 日下午馬斯克在 Twitter 表示：「若礦工能夠使用 50% 的清潔能源，並不再對環境有太大的影響，特斯拉將會恢復使用比特幣交易。」同時根據 CoinDesk 報價顯示，在這則貼文發布隔日，比特幣的價格一度漲到 41,046.77 美元，24 小時內上漲了 3.29%。

依據此實例，本專題研究透過此事件認為馬斯克在 Twitter 的言論具有深切影響力。故，欲進一步探討馬斯克的網路言論是否對於股價有所影響。

第二節 研究動機

自 2003 年起，社群媒體開始蓬勃發展，人們投入社群媒體的時間也越來越長，因此社群媒體在網路上的貼文訊息會對群眾的情緒產生影響，也間接造成了股市的漲跌。所以，如果能夠在短時間內分析預測社群網站的情感狀態，進而推測投資情緒或預測股市的走向，對於投資者來說，將可提升獲利的機率。

舉前美國總統川普為例子。川普在 Twitter 上發佈的言論時常讓人出乎意料，這是眾所皆知的事情。根據 BBC 的報導，川普曾在 2020 年一月除夕時在 Twitter 發布一則針對伊朗的言論（如圖 1-1）。想不到一則看似普通的言語，卻造成美國與伊朗爆發戰爭，也影響了全國股價。在美國及伊朗國情極度緊張的情況下，得知各國股票都有下跌的趨勢，尤其是石油的期貨下跌 4.93%，以及在全球最大黃金 ETF 道瓊財經指數基金前一交易日，黃金持有量減少 9.37 公噸至 886.81 公噸，如表 1-1 主要股價指數、債券、原物料與匯率表現所示（星展銀行，2020; BBC, 2020; Twitter, 2020）。



圖 1-1 川普針對伊朗之言論

資料來源：川普，Twitter，2020

表 1-1 主要股價指數、債券、原物料與匯率表現

名稱	收盤	漲跌	漲跌幅 (%)	名稱	收盤	漲跌	漲跌幅 (%)
道瓊工業指數	28,745.09	161.41	0.56	日經225指數	23,204.76	-370.96	-1.57
標準普爾500	3,253.05	15.87	0.49	澳洲ASX200	6,817.63	-8.81	-0.13
那斯達克指數	9,129.24	60.66	0.67	紐元指數	11,556.98	-72.88	-0.63
英國富時100	1,663.59	5.29	0.32	香港恒生指數	28,087.92	-234.14	-0.83
費城半導體指數	1,867.59	0.31	0.02	上海綜合所指數	3,066.89	-37.91	-1.22
羅素2000	44,470.91	313.10	0.71	恆生國企指數	11,079.79	-118.96	-1.06
巴西指數	116,247.00	-414.90	-0.36	台灣加權指數	11,817.10	-63.22	-0.53
道瓊歐洲600	418.36	0.69	0.17	南韓KOSPI	2,151.31	-24.23	-1.11
道瓊歐洲50	3,772.56	13.31	0.35	新加坡海峽指數	3,245.89	-1.97	-0.06
英國金融時報	7,574.93	1.08	0.01	馬來西亞指數	1,589.10	-21.94	-1.36
法國CAC指數	6,031.00	18.65	0.31	德國曼海姆指數	1,559.27	-25.96	-1.64
德國DAX指數	13,320.18	93.35	0.71	印度指數	6,225.69	-53.66	-0.85
俄羅斯指數	1,589.08	20.73	1.32	印度指數	40,817.74	-51.73	-0.13
債券 (殖利率/利率)				匯率			
美國2年期公債	1.5806	0.0402	2.61	美元指數	97.2990	0.294	0.30
美國10年期公債	1.8738	0.0561	3.09	歐元/美元	1.1105	-0.005	-0.43
美元3M Libor	1.8780	0.0000	0.00	美元/日圓	109.1200	0.680	0.63
歐元3M Libor	-0.3870	0.0000	0.00	英鎊/美元	1.3097	-0.003	-0.22
原物料				匯率			
CRB價格指數	183.38	-3.55	-1.90	澳幣/美元	0.6866	-0.000	-0.06
紐約原油期貨	59.61	-3.09	-4.93	紐幣/美元	0.6647	0.000	0.06
黃金期貨	1,560.20	-14.10	-0.90	美元/加幣	1.3039	0.004	0.27
玉米	384.25	-0.25	-0.07	美元/瑞士法郎	0.9739	0.004	0.38
小麥	552.75	2.50	0.45	美元/新幣	1.3511	0.001	0.10
黃豆	947.25	3.25	0.34	美元/人民幣	6.9462	0.001	0.01
原糖約海指數	791.00	0.00	0.00	美元/港幣	7.7754	-0.003	-0.04
				美元/台幣	30.0800	0.006	0.02

資料來源：星展銀行，每日金融及外匯資訊來源，2020

根據美國 University of California Riverside 的計算機科學與工程副教授 Vagelis Hristidis 在 2012 年對 Twitter 和金融數據之間關係的研究，在社群網站 Twitter 的推文中，證明了社群網路對於股價的影響。Vagelis Hristidis 和他的研究團隊開發了一個交易模型，利用 Twitter 反映的價格變化，在四個月的試用期內，由一億四千萬則推文和一家上市公司的股價之間的表現作為研究，當一家上市公司在 Twitter 被高度討論到時，這些輿論對該公司在股市的表現影響力就越強烈。根據研究結果，Twitter 預測表現最佳的股票下跌 2.2%，而道瓊斯工業平均指數整體下跌 4.2%。許多其他交易算法的表現也低於 Twitter 模型 (Bercovici, 2012)。

本研究採用 Twint 抓取所有推文，且透過自然語言處理 (NLTK) 將結果進行簡單的單詞統計，並將特斯拉股價作為金融數據的資料來源，來定義該推文情感狀態的漲跌，再使用 Jupyter 環境應用 Python 語言進行機器學習和自然語言的情感分析，以判斷貼文其是否會間接影響股價。

第三節 研究目的

言論往往表示了某個發言者對某事件的“意見”或是“看法”，有支持或反對，立場可能是正面、負面或者中立看法，而情感分析就是透過發文者在文章中所表達的「喜、怒、哀、樂」，來分析發文者的意見與情感對應相關資訊並加以應用。本研究目的即是分析馬斯克的網路言論是否對於特斯拉股價有所影響，用自然語言處理、情感分析、LSTM 深度學習，將資料分為訓練、驗證、測驗模型得知準確度。

第貳章 文獻回顧與探討

深度學習 (Deep Learning) 是更複雜的機器學習，同樣都是為了建立迴歸、分類模型，但唯一的差異在於深度學習使用人工類神經網路，而非其他統計模型。當中有許多不同的模型結構，通常是根據資料型態以及任務所需的功能來決定，在深度學習中最重要的兩個基本架構為卷積神經網路 (CNN, Convolution Neural Network) 以及循環神經網路 (RNN, Recurrent Neural Network)。前者被廣泛地使用在影像識別上，後者則被廣泛地使用在自然語言處理以及時間序列等問題。兩者除了資料型態不同，在結構上也有很大的不同。而本研究中的情感分析是使用 RNN 所延伸出來的 LSTM 長短期記憶作為研究主軸 (皮旭庭, 2018)。

第一節 深度學習 (Deep Learning)

深度學習是由多層神經網路 (Deep Neural Network) 發展而來。其中處理圖像資料的卷積神經網路與序列性資料的遞迴神經網路也在 1982 年後陸續被提出或改良，成為深度學習中的重要組成部分 (維基百科, 2019)。

而所謂「學習」是指在訓練模型的過程所看到的結果，告訴機器這個結果是不是好的，不是的話就要再調整函數內容，並重新產生一個新的結果。經過大量的訓練過程，最後就能找到一個最佳結果。在調整或訓練模型時，如果結果呈現出大量的錯誤答案，其實這是正常的訓練模型過程，要有錯誤才能不斷的訓練，直到精準度提高為止。這種類神經網路需要的正是訓練，它需要接收成千上萬，甚至上百萬筆資料，直到精準調整神經元的權重，讓結果幾乎接近正確答案 (DBS, 2019)。

一、循環神經網路 (RNN)

循環神經網路是一個可以儲存過去的行為記憶，用來精準的預測未來的走向，可用於預測一系列單字裡接下來的內容，還能將長度不一的資料序列送入 RNN。但 RNN 就像人的大腦一樣，一旦吸收的資訊量太大，就會比較健忘。相對來說，傳統的 RNN (如圖 2-1) 會隨著差距的擴大，而無法學習訊息之間的連接，所以若和 LSTM (如圖 2-3) 相比，會較難學習到長期的記憶 (Colah's blog, 2015)。

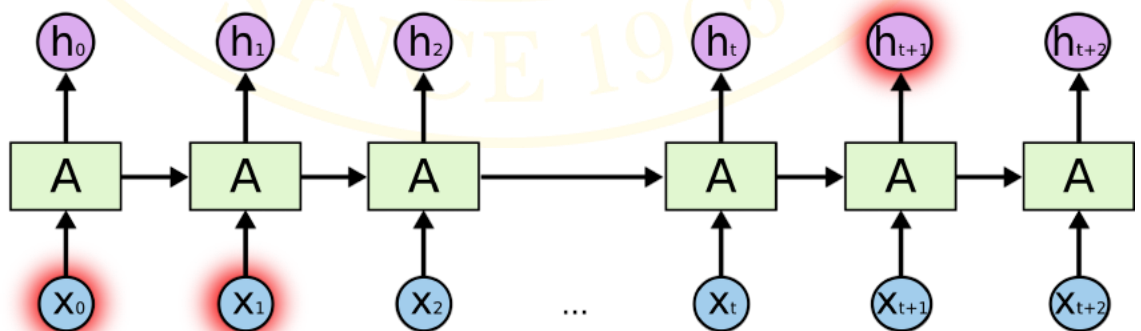
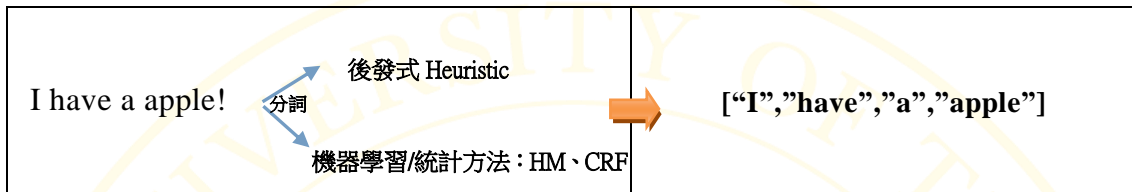


圖 2-1 傳統模型的意識圖

資料來源：Colah's blog (2015)

而所謂的雙向 RNN 是一種可以幫助一個單詞有兩種意思的判別。例如，“apple”這一個單詞，可以代表水果也可以代表品牌，會有兩種不同的意思，造成機器學習的一個錯誤。為了解決這種不確定性，因此要拆分一個句子所有的單詞（如表 2-1），以便機器做關鍵詞的判斷。而當機器在判斷關鍵詞發生疑慮時，可以進而分析“I”、“have”、“a”的句型和“apple”的結合是代表水果還是品牌。以上所舉的例子是雙向 RNN 其中一種聯繫，用於未來的學習(Martin, 2018)。

表 2-1 把長句子拆成有意義的單詞



資料來源：本研究整理

二、卷積神經網路 (CNN)

從 iT 邦幫忙的網站中可以得知卷積神經網路是模仿人類大腦的認知方式。例如，人類辨識一個圖像，會先注意到顏色鮮明的點、線、面，之後將它們構成每個不同的形狀，這種抽象化的過程就是 CNN 演算法建立模型的方式。從 NVIDIA 的網頁中可以了解到 CNN 由一個或多個隱藏層和一個輸出層所組成，這些數學運算層有助於電腦一點一點的定義影像細節，以求最後能分辨出正確의 影像，不過在辨識時常會出錯，尤其是在訓練初期。

(一)卷積層 (Convolution Layer)

卷積層主要用於對圖片做擷取特徵的處理。如圖 2-2 中間藍色的矩陣稱為卷積層，又可以叫濾波器 (filter)。不同的卷基層也有不同作用，可以幫助電腦分析影像找出影像暗區、亮區、顏色，以及高度等其它細節。

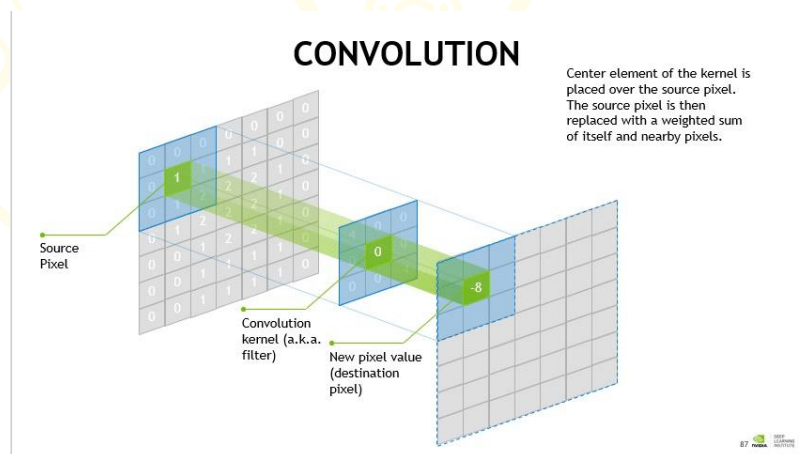


圖 2-2 卷積神經網路示意圖

資料來源：Martin, (2020)

(二)池化層 (pooling layer)

池化層是用來減少卷積層創造的容量以及減少對記憶體的需求，以置入更多卷積層，在擷取物件的特徵中取得最重要的特徵 (Martin, 2020)。

(三) 全連結層 (fully connected layer)

全連結層用於將一層裡各神經元連接到另一層裡的所有神經元。簡而言之，就是該層神經元的幾個重要特徵抽取後獲得的特徵的向量映射到下一層神經元 (Martin, 2020)。

第二節 長短期記憶模型 (LSTM)

為了解決 RNN 的健忘和誤差問題才有了長短期記憶 (Long-short-term memory, LSTM)。其特色是遞迴網路架構用三個控制閥 (Gate) 建立記憶節點 (Memory Cell)，在每次的遞迴訓練不斷的更新該節點，最後記憶的節點做出預測。這種模型屬於非線性模型，是一個相對於 RNN 更適合處理長時間序列的資料型態 (Chang, 2019)。

一、LSTM 架構

X 為輸入， h 為資料輸出，不斷地將過往資訊往下傳遞，利用過去預測得到的結果可作為未來預測的資料，以這樣的形態以此類推，形成一條長鏈，如下圖 2-3 LSTM 架構圖 1 所示 (Colah's blog, 2015)。

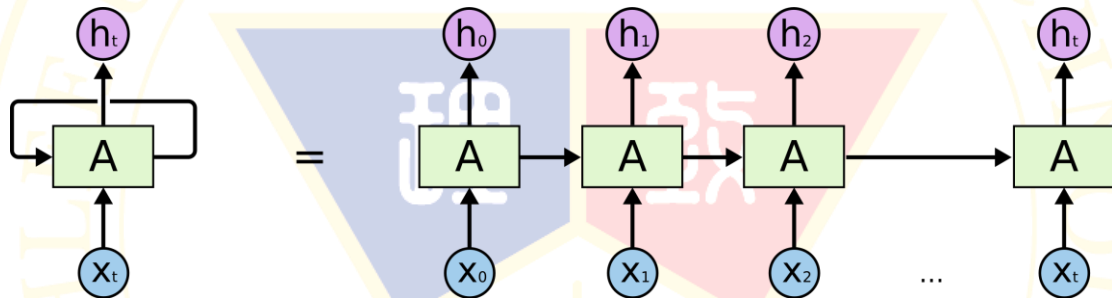


圖 2-3 LSTM 架構圖 1

資料來源：Colah's blog (2015)

在 Colah's blog (2015) 中得知，當 LSTM (如圖 2-4) 的輸入資料由 X 進入神經網路後，首先第一步會先經過四層不同的邏輯函數，為圖 2-4 中的四個黃色方塊，分別為輸入門 (input gate)、遺忘門 (forget gate)、輸出門 (output gate) 及與全連接層機活函數 (tanh)。以上四個邏輯函數可以防止循環神經網路中的梯度衰減問題，並更好的捕捉時間序列中時間差距較大的問題。

首先透過最左邊的黃色邏輯函數輸入門，用 0 跟 1 代表狀態，0 為消除，1 為保留，讓機器可以決定是否要帶入 X ，並且過濾掉不重要的資訊。保留之後會根據之前集結的資料來預測下一個單詞。此步驟的方程式即為圖 2-5 的第一條。第二步來到中間的第二個邏輯函數遺忘門以及 tanh 邏輯函數，分為兩步驟。透過圖 2-4 的中間黃色的第二個邏輯函數遺忘門來過濾更新過後的資訊，得記錄下新的資訊，然後連接第三個黃色方塊機活函數，算出一個新的向量。此步驟的方程式即為圖 2-5 的第二條及第三條。第三步是把更新過濾後的單詞，添加新的關鍵詞，最後圖 2-4 第四個黃色邏輯函數輸出門決定有多少單詞要用來輸出和最終預測的資料。此步驟的方程式為圖 2-5 的第四條、第五條及第六條。

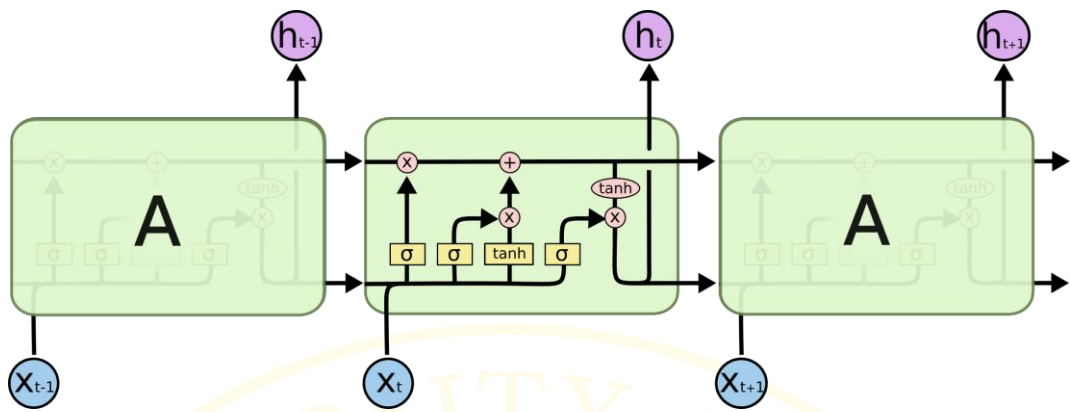


圖 2-4 LSTM 架構圖 2

資料來源：Colah's blog (2015)

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
 \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
 h_t &= o_t * \tanh(C_t)
 \end{aligned}$$

圖 2-5 計算方程式

資料來源：Colah's blog (2015)

第三節 學習預測標的

一、股票分析

(一) 股票 Stock

台灣證卷交易所網站解釋股票為有價證卷的一種，上市上櫃公司以籌措資金為目的，將公司資本分割為許多股票，而每張股票都有一個固定面額，而持有此股票可表彰投資人對發行公司之所有權，並享有公司成長或市場波動所帶來的利潤及風險。

(二) 比特幣 Bitcoin

維基百科網站中說明比特幣 (Bitcoin) 是最早推出的虛擬貨幣，又稱加密貨幣，也是目前市值排名第一的虛擬貨幣 (截至 2021 年 6 月)。比特幣的交易方式只能透過電子方式進行交易、轉移、儲存，交易的型態為代碼，沒有任何實體店面，目前這種虛擬貨幣不由中央銀行或政府機構所發行。

(三) 基本面分析

基本面分析的複雜性及邏輯性較高，適用於長期投資，在投資風險上也較低；通常會以公司的財務報表作為分析主要資料，再根據公司的競爭力及產業狀態，來評估公司的獲利能力及經營風險，這樣的分析方法比較不容易買到股價過高的股票，報酬和勝率都很高，唯一的缺點就是財報 4 個月才更新因此無法即時性的判斷。

(四) 技術面分析

技術面分析有很多指標像是黃金交叉、死亡交叉、均線、型態學、KD 指標、MACD 指標、RSI 指標、乖離率等等，複雜性及邏輯性較低，適用於短期投資，

投資風險比基本面分析來的高；主要利用圖表學或統計學作為分析資料，再根據股票的成交量、形態學做分析，例如：購買股票群眾的情緒，會反映在股價上價格的形態學，但這樣的情況很容易因為群眾對事過度的反應情緒，而導致股價與實際價值差異，而造成判斷錯誤。但這個好處是即時性高，能夠第一時間進場或出場降低虧損率。

二、估價方法

在 Mr. Marke 市場先生 (2020) 的網站上指出，過去市場上的交易量價資訊是使用圖表來觀察及預測未來價格的趨勢，以及決定投資的策略應該如何執行比較妥當。目前常見的估價方法分為三大類，分別為本益比、股價淨值比及殖利率。

(一) 本益比

本益比公式為「本益比 (P/E ratio) = 股價 (Price) ÷ 每股盈餘 (EPS)」，較低的本益比代表股票有較高的報酬率；較高的本益比代表股票有較低的報酬率，甚至容易賠錢。適用本益比估價的公司大多是高速成長的企業，因為一間企業的每股盈餘 (EPS) 處於穩定上升的時候，EPS 會比較容易預測，比起股價淨值和殖利率更適合使用本益比來估價 (慢活夫妻 George&Dewi, 2021)。

(二) 股價淨值

股價淨值比公式為「股價淨值比 = 每股價格 ÷ 每股淨值」，股價淨值比是以倍數來計算，能夠算出公司股價是淨值的多少倍，倍數越高，表示投資人就要付出更多倍的價格來買公司的資產，可以考慮買進；倍數越低，代表股價比較貴，可以考慮賣出 (Mr. Marke 市場先生, 2020)。

(三) 殖利率

殖利率公式為「股息殖利率 (%) = 每股發放現金 ÷ 每股股價」，是計算報酬率，就像銀行存款有利息一樣，把股票當成存款，把每年配發的現金股利當作利息，計算報酬率 (慢活夫妻 George&Dewi, 2021)。

第參章 研究內容與方法

第一節 馬斯克個人介紹

伊隆·馬斯克曾在台灣成立子公司，向中華民國經濟部登記負責人時使用漢名馬誼郎。他是一名出生於南非的美國企業家，同時具有南非、美國、加拿大國籍，並以 SpaceX、特斯拉汽車、PayPal，共同創辦人而聞名。馬斯克擔任 SpaceX 的執行長兼首席設計師、特斯拉汽車執行長、及 SolarCity 的董事長。他也是現代第一輛可行電動車 Tesla Roadster 的聯合設計者之一。在 2019 年富比士美國 400 富豪榜中，他以 199 億美元的資產，位列全美第 23 名富豪。在 2020 年富比士美國 400 富豪榜裡，他以 700 億美元的資產，位列為全美第 7 名富豪。2021 年 1 月 7 日，馬斯克更以超過 2,000 億美元資產成為全球首富（維基百科，2018）。



圖 3-1 馬斯克
資料來源：維基百科

第二節 研究架構

本專題抓取馬斯克推文及特斯拉股價，將蒐集的資料進行自然語言處理、情感分析和 LSTM 深度學習，並將資料分為訓練、驗證、測驗模型得知準確度（如圖 3-2）。

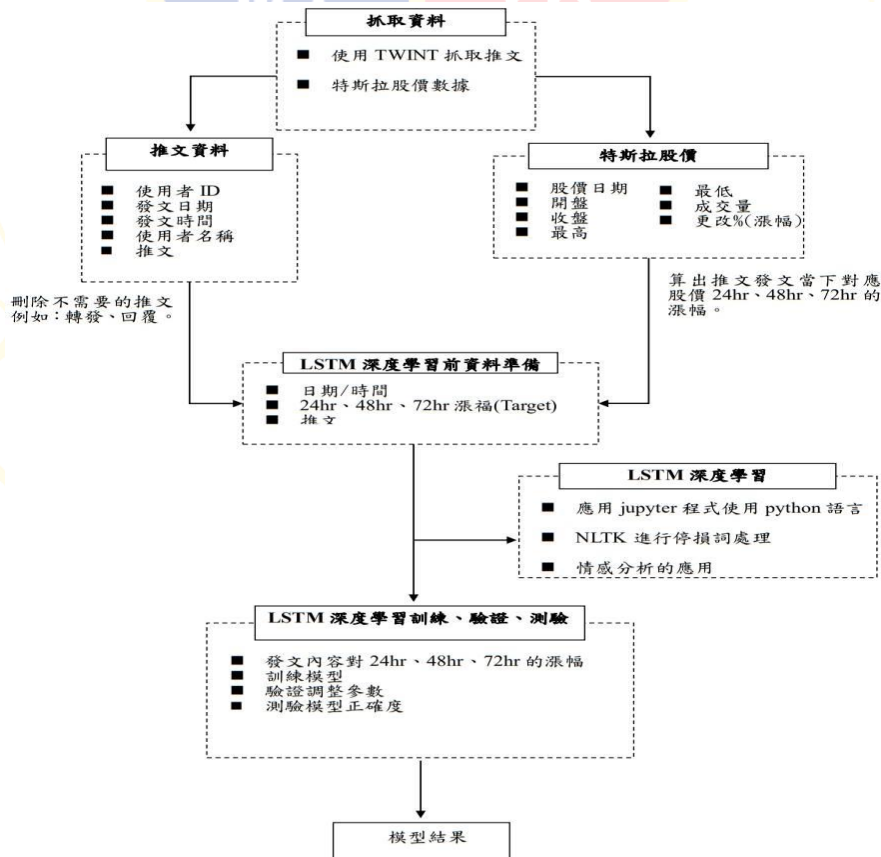


圖 3-2 研究架構圖

資料來源：本研究整理

第三節 程式使用工具

本組研究情感分析所使用之工具是使用 Jupyter Notebook 來撰寫 Python，採用 Twint 抓取馬斯克之推文，再透過 Investing.com 網站中蒐集特斯拉股票之數據，結合以上工具分析馬斯克之言論是否影響特斯拉股票之漲幅。

一、Jupyter Notebook (編寫程式分析)

Jupyter Notebook 是一個 Python 的 Web 擴充套件，能夠在瀏覽器中撰寫及執程式，主要透過撰寫 Python 程式來啟動，並利用直譯語言的特性，輸出成 HTML5，或是以 Jupyter Notebook 的形式呈現 (Summer, 2018)。

本組研究使用 Jupyter Notebook 環境來撰寫 Python，採用 Twint 抓取馬斯克之推文，再透過 Investing.com 網站中蒐集特斯拉股票之數據，結合以上工具分析馬斯克之言論是否影響特斯拉股票之漲幅。



圖 3-3 Jupyter logo 圖
資料來源：iThome (2018)

二、TWINT

Twint 是一個用 Python 撰寫的 Twitter 抓取工具。運用自動化爬蟲可以在 Twitter 上抓取幾乎所有的推文，透過 NLTK 自然語言處理將結果進行簡單的單詞統計。不像 Twitter API 限制只能抓取最後 3200 條推文 (FreeBuf.COM, 2019)。

本組研究利用 Twitter 的搜尋語法，從用戶端搜尋推文、日期、主題、標籤、圖像、影片、趨勢相關的推文，或者從推文中挑選敏感的文字訊息，例如：電子郵件、電話號碼、表情符號等。



圖 3-4 TWINT logo 圖
資料來源：FreeBuf.COM(2019)

三、抓取股市資料

本研究從 Investing.com 網站中抓取從 2015 年至 2021 年特斯拉股票數據的開盤、收盤、最高、最低、成交量及更改% (如圖 3-5)，在研究結果的部分會說明本組利用 Excel 的函數計算來對股價數據的部分進行漲跌計算整理 (如圖 4-4)。

日期	收市	開市	高	低	成交量	更改%
2021年7月2日	678.90	678.98	700.00	673.26	27.10M	0.14%
2021年7月1日	677.92	682.66	687.80	673.11	18.63M	-0.26%
2021年6月30日	679.70	680.50	692.68	678.17	18.92M	-0.16%
2021年6月29日	680.76	684.65	687.51	675.89	17.38M	-1.16%
2021年6月28日	688.72	671.66	694.34	670.58	21.63M	2.51%
2021年6月25日	671.87	689.58	693.81	668.70	32.50M	-1.17%
2021年6月24日	679.82	675.35	697.60	667.95	45.98M	3.54%
2021年6月23日	656.57	631.76	657.12	630.25	31.10M	5.27%
2021年6月22日	623.71	618.68	628.50	615.60	19.16M	0.46%
2021年6月21日	620.83	624.48	631.39	608.88	24.81M	-0.40%

圖 3-5 股價數據 1

資料來源：Investing.com, Tesla Ins

第四節 研究內容

在本專題研究中，抓取到的資料量非常龐大，這時需要透過機器學習中的三個數據集來執行，分別為訓練集、驗證集、測試集。從這些蒐集的數據中，拿取部分數據來當作測試資料，以測試模型的結果。而當測試模型結果跟驗證模型數據符合時，方能取得最佳狀態之模型，在這樣重複性的交叉驗證下，增強了本專題研究的精準度。

一、訓練集 (Training set)

訓練集就如同學習課本知識，用來訓練模型，用於模型擬合，直接參與了模型參數調整的過程。

二、驗證集 (Validation set)

驗證集就如同小考，用來知道自己哪些錯誤部份需要訂正。若要做參數調整，要把資料集在程式中切成訓練集、驗證集與測試集，然後使用訓練集訓練模型，驗證集微調參數並驗證，測試集測試模型效能。依此得知，驗證集可以評估模型效果，它的存在主要是為了調整參數，調整後可以使得模型在驗證集上的效果達至最佳。

三、測試集 (Test set)

測試集就如同期末考之最終評估。即用來測試模型正確度或預測能力。

第五節 特斯拉新聞報導

根據以下幾件新聞報導指出，馬斯克發文之後的股價漲跌幅度變化，本組蒐集了四起馬斯克的新聞報導來研究是否真的影響特斯拉股價，並加上本組所設計的程式來分析馬斯克在其他時間點的推文內容是否真的影響股價漲跌。

一、特斯拉已砸 15 億美元買比特幣、或將長期持有數位資產

根據商周 (2021) 的報導，過去馬斯克在公開場合表示支持比特幣和其他數位貨幣，去年特斯拉在 2 月 8 日宣布，特斯拉將買進價值 15 億美元的比特幣，並計畫讓顧客以比特幣作為購車的籌碼作為交易貨幣，使特斯拉將成為第一個接受

比特幣交易的汽車製造商，這樣的交易型態預估未來特斯拉可能從這筆投資獲得將近 10 億美元利潤，比起以現金交易的利潤來的還要大，當這個消息傳出後根據報導比特幣價格一度攻上 4.7 萬美元新天價，收至 4.3 萬美元，上漲 5,064 美元，漲幅達 13.09%。

二、馬斯克口無遮攔，特斯拉「熊出沒」！

根據華盛頓郵報（2021）的報導，美國白宮將提議對美國的億萬富豪加徵新稅，這項政策在美國未來十年可減少將近 3600 億美元的赤字，身為世界首富馬斯克應當也會成為支付稅務的其中一人，根據加州大學的經濟學家表示，關於這項政策馬斯克將於每年額外支付 500 億美元的稅款。此消息一發布後馬斯克回文以諷刺的形式來暗示可能會拋售更多的股票，此則貼文發出後讓網友紛紛感到不滿，認為馬斯克這番言論有失風度，這樣的一個發言導致了特斯拉股價持續下滑，一度來到每股 978.60 美元的單日低點，較 11 月 4 日觸及的 1,229.91 美元的歷史高點回落逾 20%，一度跌入熊市。

三、推特民意調查支持馬斯克出售特斯拉 10% 股份

根據 BBC（2021）的報導，特斯拉（Tesla）執行長馬斯克（Elon Musk）於 2021 年 11 月 6 日在 Twitter 發起投票，表示他已賣出「足夠的股票」，來兌現他出售特斯拉 10% 持股的承諾，有超過 350 萬的推特用戶參與，將近 58% 的用戶支持出售特斯拉股票，而馬斯克也同意履行投票結果。在結果出來的幾天後，特斯拉股價大跌 4.8%。

四、股價太高了，特斯拉應聲暴跌 10%

根據鉅亨（2020）的報導，馬斯克在 2020 年 5 月 1 日於推特上表示：「特斯拉股價太高了」還緊接著發布其他條推文，內容包括引述部分美國國歌歌詞、詩人 Dylan Thomas 的詩句，馬斯克更提到將出售他所有「有形」財產，這些推文讓特斯拉股市少了 130 億美元，但是馬斯克卻表示自己完全是在開玩笑，沒有其他的意思。

第肆章 研究結果與設計

第一節 蒐集資料

一、抓取馬斯克之推文

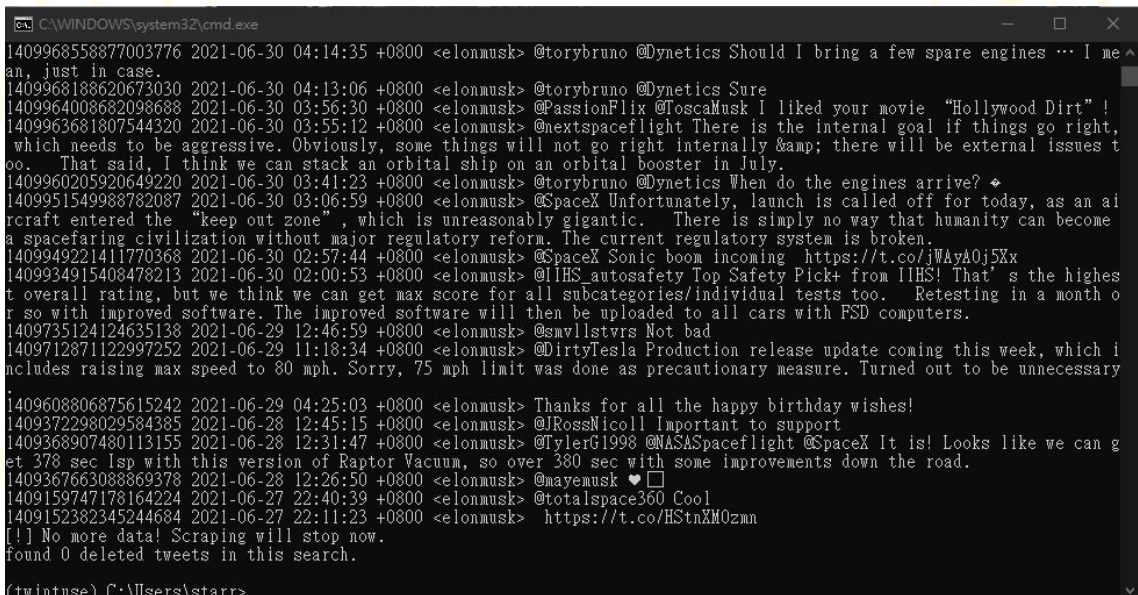
本研究採用的方式為 Twint 程式，該方法是利用一種使用 Python 編寫的高級 Twitter 抓取工具，直接從 Twitter 中抓取推文，裡面的推文內容會出現主題標籤 (#Hashtag)、網址、表情符號 (emjio) 等多餘資訊，所以本組使用 Twint 來抓取馬斯克的推文，篩選掉不必要的網址或轉貼文章等，讓資料庫只出現我們需要分析的貼文，如下圖 4-1 即圖 4-2 所示。



```
C:\WINDOWS\system32\cmd.exe
(twintuse) C:\Users\starr>twint -u elonmusk --since 2021-06-27 -o musk.csv --csv
```

圖 4-1 程式碼 1

資料來源：本研究整理



```
C:\WINDOWS\system32\cmd.exe
1409968558877003776 2021-06-30 04:14:35 +0800 <elonmusk> @torybruno @Dynetics Should I bring a few spare engines ... I me
an, just in case.
1409968188620673030 2021-06-30 04:13:06 +0800 <elonmusk> @torybruno @Dynetics Sure
1409964008682098688 2021-06-30 03:56:30 +0800 <elonmusk> @PassionFlix @ToscaMusk I liked your movie "Hollywood Dirt" !
1409963681807544320 2021-06-30 03:55:12 +0800 <elonmusk> @nextspaceflight There is the internal goal if things go right,
which needs to be aggressive. Obviously, some things will not go right internally & there will be external issues t
oo. That said, I think we can stack an orbital ship on an orbital booster in July.
1409960205920649220 2021-06-30 03:41:23 +0800 <elonmusk> @torybruno @Dynetics When do the engines arrive? ✨
1409951549988782087 2021-06-30 03:06:59 +0800 <elonmusk> @SpaceX Unfortunately, launch is called off for today, as an ai
rcraft entered the "keep out zone", which is unreasonably gigantic. There is simply no way that humanity can become
a spacefaring civilization without major regulatory reform. The current regulatory system is broken.
1409949221411770368 2021-06-30 02:57:44 +0800 <elonmusk> @SpaceX Sonic boom incoming https://t.co/jW4yA0j5Xx
1409934915408478213 2021-06-30 02:00:53 +0800 <elonmusk> @IHS_autosafety Top Safety Pick+ from IHS! That' s the highes
t overall rating, but we think we can get max score for all subcategories/individual tests too. Retesting in a month o
r so with improved software. The improved software will then be uploaded to all cars with FSD computers.
1409735124124635138 2021-06-29 12:46:59 +0800 <elonmusk> @smvllstvrz Not bad
1409712871122997252 2021-06-29 11:18:34 +0800 <elonmusk> @DirtyTesla Production release update coming this week, which i
ncludes raising max speed to 80 mph. Sorry, 75 mph limit was done as precautionary measure. Turned out to be unnecessary
1409608806875615242 2021-06-29 04:25:03 +0800 <elonmusk> Thanks for all the happy birthday wishes!
1409372298029584385 2021-06-28 12:45:15 +0800 <elonmusk> @IRossNicoll Important to support
1409368907480113155 2021-06-28 12:31:47 +0800 <elonmusk> @TylerG1998 @NASASpaceflight @SpaceX It is! Looks like we can g
et 378 sec lsp with this version of Raptor Vacuum, so over 380 sec with some improvements down the road.
1409367663088869378 2021-06-28 12:26:50 +0800 <elonmusk> @mayemusk ❤️👍
1409159747178164224 2021-06-27 22:40:39 +0800 <elonmusk> @totalspace360 Cool
1409152382345244684 2021-06-27 22:11:23 +0800 <elonmusk> https://t.co/HStnXMOzmn
[!] No more data! Scraping will stop now.
found 0 deleted tweets in this search.
(twintuse) C:\Users\starr>
```

圖 4-2 程式碼 2

資料來源：本研究整理

二、抓取股市資料

本研究從 Investing.com 網站中抓取特斯拉股票的數據及更改% (如圖 4-3)。

特斯拉 678.90 +0.98 (+0.14%)

日期	收市	開市	高	低	成交量	更改%
2021年6月30日	679.70	680.50	692.68	678.17	18.92M	-0.16%
2021年6月29日	680.76	684.65	687.51	675.89	17.38M	-1.16%
2021年6月28日	688.72	671.66	694.34	670.58	21.63M	2.51%
2021年6月25日	671.87	689.58	693.81	668.70	32.50M	-1.17%
2021年6月24日	679.82	675.35	697.60	667.95	45.98M	3.54%
2021年6月23日	656.57	631.76	657.12	630.25	31.10M	5.27%
2021年6月22日	623.71	618.68	628.50	615.60	19.16M	0.46%
2021年6月21日	620.83	624.48	631.39	608.88	24.81M	-0.40%
2021年6月18日	623.31	613.37	628.35	611.80	24.56M	1.09%
2021年6月17日	616.60	602.29	621.45	601.47	22.70M	1.94%
2021年6月16日	604.87	597.10	608.50	593.67	22.14M	0.92%
2021年6月15日	599.36	616.42	616.50	598.43	17.76M	-2.97%
2021年6月14日	617.69	612.19	624.83	609.30	20.42M	1.28%
2021年6月11日	609.89	610.23	612.56	601.52	16.21M	-0.04%
2021年6月10日	610.12	603.91	616.29	600.87	23.92M	1.89%
2021年6月9日	598.78	603.05	611.48	597.80	16.58M	-0.80%
2021年6月8日	603.59	623.01	623.01	595.60	26.05M	-0.25%
2021年6月7日	605.13	590.55	608.11	582.88	22.54M	1.01%
2021年6月4日	599.05	579.71	600.61	577.20	24.04M	4.58%
2021年6月3日	572.84	602.12	604.53	571.25	30.11M	-3.33%
2021年6月2日	605.12	620.63	623.09	599.14	23.30M	-3.01%
2021年6月1日	623.90	627.80	633.80	620.55	17.92M	-0.21%
最高: 697.60	最低: 571.25	差價: 126.35	平均: 626.92	更改%: 8.71		

圖 4-3 股價數據 1

資料來源：Investing.com , Tesla Ins

根據馬斯克發佈推文後 24 小時內的股價波動去計算是否推文會影響股票的漲跌。本組採用的計算方式為： $(\text{現價} - \text{上一交易日的收盤價}) \div \text{上一交易日的收盤價} \times 100\%$ ，將此結果的高低值分成三個等級作為漲跌依據，漲跌在 0 到 0.5% 以內的數字設為 2；漲跌在 0.5% 以上的數字設為 4；0 以下的數字設為 0，若遇假日無數據時，則採用隔週星期一的數據計算。

如圖 4-4 所示，在 2021/6/21 發佈之推文在 24 小時 (target 欄位) 內所影響的漲跌為 0；在 48 小時 (48target 欄位) 內影響的漲跌為 2；在 72 小時 (72target 欄位) 內影響的漲跌為 4。此方法是利用情感分析的概念，把情緒的正面、負面情緒，改成用漲與跌表達，例如 4 代表情緒名稱是正面，而本組經過討論之後把代表正面情緒的 4 修改成漲。

target	48target	72target	id	date	time	timezone	username	tweet
4	4	4	1.40856E+18	2021/6/26	06:58:58	800	elonmusk	All 72 orbital planes activate in
4	4	0	1.40856E+18	2021/6/26	06:51:28	800	elonmusk	Starlink simultaneously active users
0	4	0	1.40838E+18	2021/6/25	19:03:04	800	elonmusk	My Shiba Inu will be named Floki
0	4	0	1.40838E+18	2021/6/25	18:42:24	800	elonmusk	Timing is everything
0	4	0	1.40825E+18	2021/6/25	10:11:52	800	elonmusk	"That's not funny!" - Bitcoin
0	4	0	1.40825E+18	2021/6/25	10:10:31	800	elonmusk	How many Bitcoin maxis does it take to screw in a lightbulb?
4	0	4	1.40787E+18	2021/6/24	09:10:10	800	elonmusk	Shout out to the awesome Tesla global team. Thanks for working so hard to make Tesla successful!
2	4	4	1.40705E+18	2021/6/22	02:59:34	800	elonmusk	Looking at holding Tesla AI Day in
0	2	4	1.40671E+18	2021/6/21	04:19:20	800	elonmusk	Happy Father's Day!

圖 4-4 整理數據 1

資料來源：本研究整理

第二節 LSTM-sentiment-analysis

一、匯入資料

首先利用 read_excel 函數把收集好的資料庫 (xlsx 檔) 匯入 Jupyter，接著因後續切割資料需使用 csv 檔，故先將資料庫轉檔 (如圖 4-5)。


```

In [114]: data=pd.read_excel("TSLA_musk0628.xlsx",header= 0)

In [115]: data
0  跌  1408380216653839872  2021-06-25  19:03:04  800  elonmusk  My Shiba Inu will be named Floki
1  跌  1408375014261390080  2021-06-25  18:42:24  800  elonmusk  Timing is everything
2  跌  1408246534093910016  2021-06-25  10:11:52  800  elonmusk  "That's not funny!" – Bitcoin
3  跌  1408246196603340032  2021-06-25  10:10:31  800  elonmusk  How many Bitcoin maxis does it take to screw i...
4  漲  1407868620680689920  2021-06-24  09:10:10  800  elonmusk  Shout out to the awesome Tesla global team. Th...
...  ...  ...  ...  ...  ...  ...  ...
2959  漲  552437001291841024  2015-01-06  20:10:19  800  <elonmusk>  Need to investigate the upper stage Z actuator...
2960  漲  552279321491275008  2015-01-06  09:43:45  800  <elonmusk>  Taking questions now at http://t.co/CpOyWeAKOG
2961  漲  552256055607312000  2015-01-06  08:11:18  800  <elonmusk>  Background on our attempt to land Falcon 9 roc...
2962  漲  552182273865699008  2015-01-06  03:18:07  800  <elonmusk>  Drone spaceport ship heads to its hold positio...
2963  漲  552153757338065984  2015-01-06  01:24:48  800  <elonmusk>  Ask me anything at 9pm Florida time (focused o...

2964 rows x 7 columns

In [116]: def xlsx_to_csv_pd(): # musk06028.xlsx 轉換成csv檔
data_xls = pd.read_excel('TSLA_musk0628.xlsx', header=0)
data_xls.to_csv('csv0628.csv', encoding='utf-8')
return data_xls

if __name__ == '__main__':
data_xls_csv= pd.read_excel('TSLA_musk0628.xlsx', index_col=0)
xlsx_to_csv_pd()

```

圖 4-5 程式碼 3

資料來源：本研究整理

二、資料移除

因抓取下來的推文內容還未經過處理，像是轉發文章、回覆文章留言內容，又或者是網址這類不具有文字意義的字串，都必須先將它們移除。接著因後續分析只需使用到 target 跟 tweet 欄位，故抓取這兩個欄位即可，並且使用自然語言處理工具包（Natural Language Tool Kit, NLTK）移除停用詞（Stop Words），如圖 4-6 及圖 4-7 所示。

```

In [123]: # # 刪除非字母數字字符
musk_sentiment["tweet"] = [re.sub(r'https?:\/\/\S+', '', str(x)) for x in musk_sentiment["tweet"]]
# # 刪除轉發
musk_sentiment= musk_sentiment[~musk_sentiment["tweet"].str.startswith("RT @")]
musk_sentiment["tweet"] = musk_sentiment["tweet"].replace("", np.nan)
musk_sentiment= musk_sentiment.dropna(subset = ["tweet"])

musk_sentiment=musk_sentiment.drop(columns=['Unnamed: 0'])

In [127]: musktweet = musk_sentiment[['target', 'tweet']]
musktweet.head(10)

Out[127]:
   target  tweet
0  跌  My Shiba Inu named Floki
1  跌  Timing everything
2  跌  "That's funny!" – Bitcoin
3  跌  How many Bitcoin maxis take screw lightbul?
4  漲  Shout awesome Tesla global team. Thanks workin...
5  平盤  Looking holding Tesla AI Day
6  跌  Happy Father's Day!
7  漲  Real pic 2 ships next Starbase Tiki Bar right
8  漲  Let's skip original video games & movies &...
9  跌  From top Starbase high bay

```

圖 4-6 訓練執行結果 1

資料來源：本研究整理

```
In [128]: from nltk.corpus import stopwords
from nltk.stem import SnowballStemmer #SnowballStemmer為詞幹提取器介面
nltk.download("stopwords")
stop_words = stopwords.words("english") #這裡設定稍後取用 English 的停用詞語料庫
stemmer = SnowballStemmer("english")
def preprocess(text, stem=False):
    # Remove link, user and special characters
    tokens = []
    for token in text.split():
        if token not in stop_words:
            if stem:
                tokens.append(stemmer.stem(token))
            else:
                tokens.append(token)
    return " ".join(tokens)

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\user\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

圖 4-7 訓練執行結果 2

資料來源：本研究整理

如圖 4-8 所示，本組經過程式的推文處理後，就沒有再出現轉發的文字內容。

```
In [29]: musk_sentiment.tail(40)
Out[29]:
```

	Unnamed: 0	target	id	date	time	timezone	username	tweet
2924	2924	平蓋	564562686478806976	2015-02-09	07:13:28	800	<elonmusk>	Air Force tracking radar went down. Launch pos...
2925	2925	平蓋	564509965612633984	2015-02-09	03:43:58	800	<elonmusk>	Rocket reentry will be much tougher this time ...
2926	2926	平蓋	564493608351313024	2015-02-09	02:38:58	800	<elonmusk>	Launching our 1st deep space mission today. He...
2927	2927	跌	561002248827052032	2015-01-30	11:25:33	800	<elonmusk>	If you are curious about the P85D, you can sch...
2928	2928	跌	560909571691379968	2015-01-30	05:17:17	800	<elonmusk>	Painting the name on the dronship ... http/...

```
In [31]: musk_sentiment.tail(40)
Out[31]:
```

	target	id	date	time	timezone	username	tweet
2924	平蓋	564562686478806976	2015-02-09	07:13:28	800	<elonmusk>	Air Force tracking radar went down. Launch pos...
2925	平蓋	564509965612633984	2015-02-09	03:43:58	800	<elonmusk>	Rocket reentry will be much tougher this time ...
2926	平蓋	564493608351313024	2015-02-09	02:38:58	800	<elonmusk>	Launching our 1st deep space mission today. He...
2927	跌	561002248827052032	2015-01-30	11:25:33	800	<elonmusk>	If you are curious about the P85D, you can sch...
2928	跌	560909571691379968	2015-01-30	05:17:17	800	<elonmusk>	Painting the name on the dronship ... http/...

圖 4-8 訓練執行結果 3

資料來源：本研究整理

三、準備數據

我們使用 tensorflow.keras 提供的 tokenizer，將文本中的每一個文字賦予一個獨有的數值，進而建立起一個字典，如圖 4-9 所示。

```
In [53]: from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences

tweet = musktweet.tweet.values
tokenizer = Tokenizer(num_words=5000)
tokenizer.fit_on_texts(tweet)
vocab_size = len(tokenizer.word_index) + 1
encoded_docs = tokenizer.texts_to_sequences(tweet)
padded_sequence = pad_sequences(encoded_docs, maxlen=300)

In [54]: print(tweet[0])
print(encoded_docs[0])

My Shiba Inu will be named Floki
[43, 1921, 1922, 11, 12, 823, 2909]
```

圖 4-9 程式碼 4

資料來源：本研究整理

第三節 運用深度學習訓練模型

一、資料分割

我們將資料切分成訓練集 (Training Data) 以及測試集 (Testing Data)，並透過 `train_test_split` 這個函數來做資料分割。保留較多的資料 (80%) 作為訓練集，其餘的資料做為測試集 (如圖 4-10)。

```
In [133]: from sklearn.model_selection import train_test_split

df_train, df_test = train_test_split(musktweet, test_size=0.2, random_state=1)
print("TRAIN size:", len(df_train))
print("TEST size:", len(df_test))

df_train

TRAIN size: 2186
TEST size: 547

Out[133]:
```

	target	tweet
202	跌	Cryptocurrency explained
1784	跌	It's real
653	跌	2019 seems quaint & long ago
625	跌	Gave proof night flag still
1562	漲	Tesla acquired trucking capacity ensure Model ...
...
1018	跌	Tesla commercial solar ordering live
948	漲	Homes solar sell 4.1% more, according Zillow
1160	跌	On Mars, I like drink coffee mug
242	跌	Doge
1120	跌	New @Hyperloop top speed 463 km/hll

2186 rows x 2 columns

圖 4-10 程式碼 5

資料來源：本研究整理

二、訓練集準備

然而發現我們並未把 Target 做資料切割與轉換成數字矩陣，無法和 Tweet 訓練，因為 Tweet 已轉換成數字矩陣，導致與 Target 不相容。根據此問題點，我們著手研究改善方法並加以修正。

在此階段本組過濾掉分類為「平盤」之文章，篩選出分類為「漲」與「跌」之文章。使用 `factorize` 函數返回一個數字類別索引，指數 0 是跌的情緒，指數 1 是漲的情緒，此方法能讓 Target 值轉換成數字矩陣，以利後續 Target 與 Tweet 能相容且進行訓練。

因推文的文字長短不同，所以利用 `pad_sequence` 函數使陣列長度等長，如圖 4-11 所示。

```
In [129]: musktweet = musktweet[musktweet['target'] != "平盤"]
musktweet.head(10)

Out[129]:
```

	target	tweet
0	跌	My Shiba Inu named Floki
1	跌	Timing everything
2	跌	"That's funny!" - Bitcoin
3	跌	How many Bitcoin maxis take screw lightbulb?
4	漲	Shout awesome Tesla global team. Thanks workin...
6	跌	Happy Father's Day!
7	漲	Real pic 2 ships next Starbase Tiki Bar right
8	漲	Let's skip original video games & movies &...
9	跌	From top Starbase high bay
10	跌	Stacking Super Heavy Aft Section

```
In [130]: sentiment_label = musktweet.target.factorize()
sentiment_label

Out[130]: (array([0, 0, 0, ..., 1, 1, 1], dtype=int64),
Index(['跌', '漲'], dtype='object'))
```

圖 4-11 程式碼 6

資料來源：本研究整理

三、學習調整

如圖 4-11 顯示，loss 代表訓練集損失值，accuracy 代表訓練集準確率，所以 loss 與 accuracy 為訓練模型，而 val_loss 是驗證集損失值，val_accuracy 代表驗證集準確率，所以 val_loss 與 val_accuracy 為測驗模型。同時可以觀察訓練模型中的 loss 及測驗模型中的 val_loss 來判斷此模型是處於何種學習狀態。

在本組的研究過程中，我們可以看見模型的訓練結果，如圖 4-12 所見，loss 不斷下降，val_loss 趨於不變，本組推測此模型可能呈現過擬合（overfitting）之狀態。

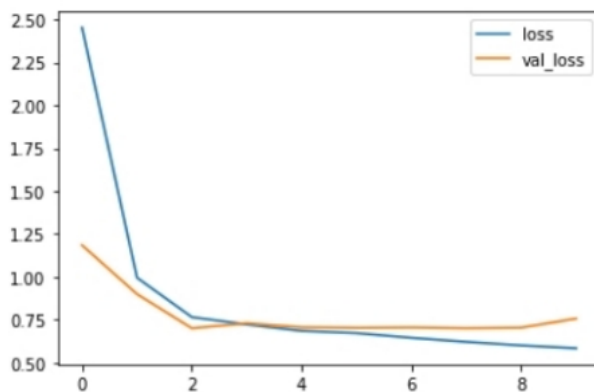


圖 4-12 訓練執行結果 3

資料來源：本研究整理

為了調整以上之問題，我們首先定義狀態，並且初始化為 0。接下來使用 embedding layer 來對句子進行編碼及分配不同的索引。之後，放至 LSTM，並放到一個全連接層（Dense）模型進行分類。

我們使用添加 dropout 正規化，利用不同的神經網路訓練之後，求其輸出的平均值，在每個樣本中隨機刪除不同部分的神經元，dropout=0.2 這裡我們使用比例是 20%，如圖 4-13 所示。

```

In [206]: batchsz = 512
data = tf.data.Dataset.from_tensor_slices((x_train,y_train))
data = data.shuffle(10000).batch(batchsz,drop_remainder=True)

data_test = tf.data.Dataset.from_tensor_slices((x_test,y_test))
data_test = data_test.batch(batchsz,drop_remainder=True)

In [207]: units=128
batchsz =512
total_words = vocab_size
embedding_len=128

class Simple_LSTM(keras.Model):
    def __init__(self,units):
        super(Simple_LSTM, self).__init__()
        self.state_0 = [tf.zeros([batchsz,units])]
        self.embedding = layers.Embedding(total_words,embedding_len,input_length=300)
        self.layer_1 = layers.SimpleRNNCell(units,dropout=0.2)
        self.out = layers.Dense(1)

    def call(self, inputs, training=None):
        x = self.embedding(inputs)
        state_0 = self.state_0
        for word in tf.unstack(x,axis=1):
            out,state_1 = self.layer_1(word,state_0)
            state_0 = state_1
        x = self.out(out)
        prob = tf.sigmoid(x)
        return prob

```

圖 4-13 程式碼 7

資料來源：本研究整理

第四節 研究結果

透過以上情況，我們可以得知本模型 loss 及 val_loss 不斷下降，說明損失值逐漸減少，模型正在學習，也可以看見 accuracy 及 val_accuracy 不斷上升，說明模型準確率逐漸提升，此種狀況為最好的學習狀態。

最初將 epochs 設為 10、及 20（圖 4-14），但訓練出來的 loss、val_loss 最終沒有密合，這代表並沒有學習完成。

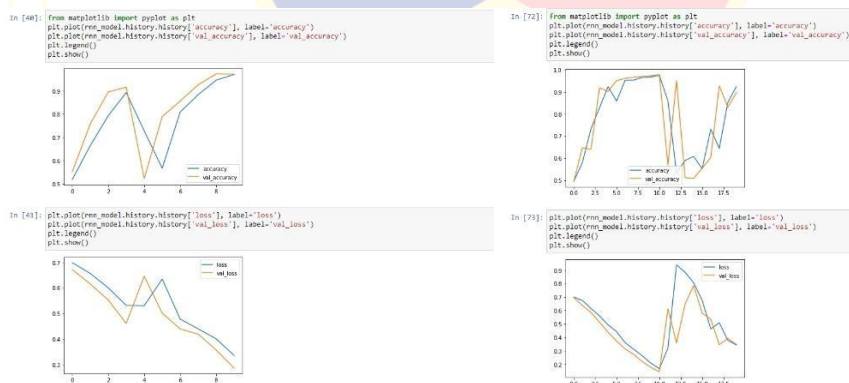


圖 4-14 程式碼 8

資料來源：本研究整理

後來我們將 epochs 設為 30，可以透過圖 4-15 觀察到，LSTM 模型的 accuracy 及 val_accuracy 不斷上升，說明模型準確率逐漸提升。loss 及 val_loss 不斷下降，說明模型正在學習且學習率最為良好損失值也逐漸減少。

```

In [31]: epochs = 30
rnn_model = Simple_LSTM(units)
rnn_model.compile(optimizer=keras.optimizers.Adam(1e-3), loss=tf.losses.BinaryCrossentropy(), metrics=['accuracy'])
rnn_model.fit(data, epochs=epochs, validation_data = data_test)

Epoch 25/30
4/4 [=====] - 2s 429ms/step - loss: 0.0365 - accuracy: 0.9937 - val_loss: 0.0329 - val_accuracy: 0.9956
Epoch 26/30
4/4 [=====] - 2s 489ms/step - loss: 0.0314 - accuracy: 0.9956 - val_loss: 0.0303 - val_accuracy: 0.9946
Epoch 27/30
4/4 [=====] - 2s 480ms/step - loss: 0.0296 - accuracy: 0.9966 - val_loss: 0.0280 - val_accuracy: 0.9956
Epoch 28/30
4/4 [=====] - 2s 403ms/step - loss: 0.0280 - accuracy: 0.9937 - val_loss: 0.0261 - val_accuracy: 0.9956
Epoch 29/30
4/4 [=====] - 2s 491ms/step - loss: 0.0263 - accuracy: 0.9956 - val_loss: 0.0244 - val_accuracy: 0.9956
Epoch 30/30
4/4 [=====] - 2s 581ms/step - loss: 0.0238 - accuracy: 0.9956 - val_loss: 0.0228 - val_accuracy: 0.9956

Out[31]: <tensorflow.python.keras.callbacks.History at 0x1c69a467370>

```

圖 4-15 程式碼 9

資料來源：本研究整理

該模型經過 30 個 Epoch 的訓練，由圖 4-16 及圖 4-17 可見，accuracy 及 val_accuracy 和 loss 及 val_loss 兩者之線條，最後的差異都不大，且近乎相同，說明測驗及驗證的準確率都有達到標準，且高達 99%。

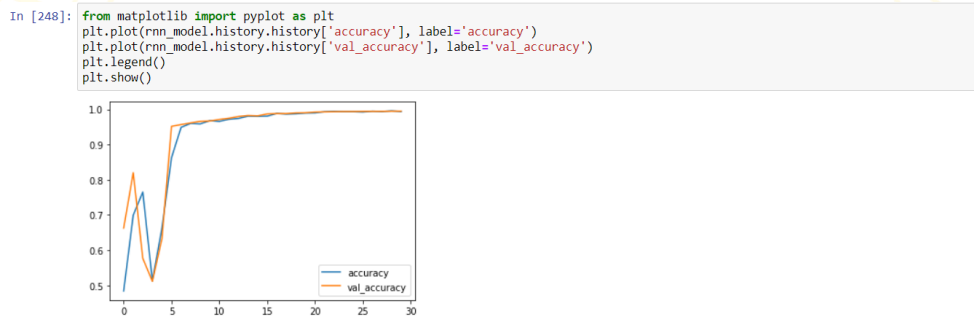


圖 4-16 程式碼 10

資料來源：本研究整理



圖 4-17 程式碼 11

資料來源：本研究整理

上述所做的分析皆為研究推文發佈後 24 小時之股價影響，我們也針對 48 小時與 72 小時做 LSTM 模型，最後利用 Test_accuracy (測驗準確度) 去做對比，Test_accuracy 分別為 24 小時 0.9956、48 小時 0.9946 以及 72 小時 0.9975 (圖 4-18)。

第五章 結論與未來展望

第一節 結論

因為社群網路的發展導致越來越多人重視一個人所說的話會影響到什麼，尤其是那些有權有利的人說出來的話往往會影響經濟變動。最明顯可見的例子如美國總統川普和特斯拉執行長馬斯克。他們的推文影響力甚至可以撼動股市。故藉由本研究結果可以分析 2015 年至 2021 年馬斯克在推特說的話會對哪些市場上的汽車股市有所影響，並且可以有利於特斯拉提前做出應對以及股市接下來的變化。

一、不同時間所造成的股價影響

本組在情感分析中，先利用 Twint 抓取馬斯克的推文，再用 Stopwords 對推文做字詞處理以及拿掉不必要的符號，最後再用 Tokenizer 將本研究的推文向量化，把推文轉換為序列，以利模型進行分詞的處理。整理完之後再將推文與 Target 去做比對，查看馬斯克的推文對股價漲跌的影響，為了確保準確性，本組又將 Target 分成了三個時段去做比對，分別是推文發布的當下所影響 24 小時、48 小時、72 小時的股價漲跌，再把這三個時段的 Target 拿去 test_accuracy 做比對來測試準確度，跑過程式後即可得知這三個時段的準確率分別為多少。依本研究的結果得知，在推文發布後的 72 小時內影響股票漲跌與我們的訓練最相符合，其準確率高達 0.9975，而我們也可以推測有時訊息傳播速度較慢，股市可能到兩天、或三天後才開始有變化，不同的時間長短影響股票的漲跌也不盡相同。

二、LSTM 模型改善

在本組做機器學習的過程中，因模型曾呈現學習不佳的狀態，程式中使用的 Batchsz 較大，導致所需的記憶體較大，原有的 Epoch 訓練次數相對 Batchsz 較小，所以本組調整了 Epoch，當訓練的次數增加，即會減少學習不佳的狀態。除了修正 Epoch 的訓練次數外，同時也對資料做了轉換，因為本研究的 Tweet 已經轉換成數字矩陣，但 Target 的資料還未做轉換，導致兩個資料並不相容，所以使用了 factorize 函數，用此方法讓 Target 轉換成數字矩陣，這樣 Target 與 Tweet 的資料就能相容的進行訓練，以此來改善學習不佳的狀態。

三、研究建議

在情感分析的研究過程中，曾發生過 loss 下降但 accuracy 不變的情形。此現象的可能原因有四點。第一點，數據集合樣本各類別數量差距太大；第二點，訓練集和數據集是手動劃分的話，就要改為代碼自動劃分，減少網路層數，多設置 epoch；第三點，訓練模型不適用或者數據不適當，建議調整參數或是改算法；第四點，資料格式未轉為相同。往後做情感分析時應將以上這些因素納入考量。

第二節 未來展望

本研究使用情感分析馬斯克的推文，用了三種模型集來呈現情感分析過程，分為訓練模型、驗證模型、測試模型。此三種模型是相輔相成的，透過保

留訓練模型的百分之八十來對剩下百分之二十的測試模型做預測，在經過 30 次的訓練及測試中，準確度不斷的提高到 0.9。

因為股市的上下波動及影響因素很多，本組考慮未來將其他汽車股市加入模型中做為參考值之一，觀察是否也會影響到此汽車股市，也期望本組研究能應用在預測未來股價上。模型同時也可導入多面向分析之長期模型，並使用算法融合思想提高情感分析的精準度。本研究又或可往更多面向的股價來套入練習模型，例如：油價、現金股價、電商股價、科技股價，再利用情感分析不同語句跟隱藏式情感詞彙之間的差異，觀察上下則推文的字面意義，以解決情感分析訊息不足的問題，並且獲得更精確的預測結果。



參考文獻

一、中文部分

1. AI StockBoy (2019), 學會用機器學習預測股價 — 完整流程教學與實作, 2021年4月17日, 擷取自:
<https://medium.com/ai%E8%82%A1%E4%BB%94/%E5%AD%B8%E6%9C%83%E7%94%A8%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92%E9%A0%90%E6%B8%AC%E8%82%A1%E5%83%B9-%E5%AE%8C%E6%95%B4%E6%B5%81%E7%A8%8B%E6%95%99%E5%AD%B8%E8%88%87%E5%AF%A6%E4%BD%9C-b057e7343ca4>
2. AK Share (無日期), AKShare 股票数据, 2021年1月5日, 擷取自:
https://www.akshare.xyz/zh_CN/latest/data/stock/stock.html
3. Andy Chen (2018), 到底什麼是電動車? EV、BEV、PHEV 又各是什麼呢?, 2021年1月10日, 擷取自:
<https://www.chargesmith.com/ev-pev-phev-compare/>
4. A soul (2017), GitHub 台灣股票即時爬蟲, 2021年1月5日, 擷取自:
<https://github.com/Asoul/tsrtc>
5. BBC (2020), 美國大選: 特朗普當總統四年來改變了什麼, 2021年1月13日, 擷取自:
<https://www.bbc.com/zhongwen/trad/world-54688655>
6. Chang, T. (2019), 淺談遞歸神經網路 (RNN) 與長短期記憶模型 (LSTM), 2021年07月05日, 擷取自:
<https://pse.is/3j4ycc>
7. Dan (2019), LSTM - 股價預測 (Data: 台灣 50), 2021年5月3日, 擷取自:
<https://ithelp.ithome.com.tw/m/articles/10224356>
8. DBS (2019), AI 時代大揭密 | 何謂「深度學習」?, 2021年1月20日, 擷取自:
<https://medium.com/codingbar/ai-%E6%99%82%E4%BB%A3%E5%A4%A7%E6%8F%AD%E5%AF%86-%E4%BD%95%E8%AC%82-%E6%B7%B1%E5%BA%A6%E5%AD%B8%E7%BF%92-c5e81f240a5b>
9. FreeBuf.COM (2019), TWINT: 一款 Twitter 信息爬取工具- FreeBuf 網路安全行業門戶, 2021年1月20日, 擷取自:
<https://www.freebuf.com/sectool/212601.html>
10. IDC 國際數據庫 (無日期), 關於 IDC (國際數據資訊有限公司), 2021年1月10日, 擷取自:
<https://www.idc.com/tw>
11. IDC: 未來五年中國新能源汽車市場將強勁增長, 2021年7月5日, 擷取自:
<https://new.qq.com/omn/20201204/20201204A03SFG00.html>
12. iThome (2018), LSTM - 股價預測, 2021年3月9日, 擷取自:
<https://ithelp.ithome.com.tw/articles/10224356>
13. itread01.com (2018), 使用 Python 的自然語言工具包 (NLTK) 對 Reddit 新聞標題進行情感分析, 2021年7月8日, 擷取自:
<https://www.itread01.com/content/1544761110.html>
14. IT人 (2019), NLP 入門 (十) 使用 LSTM 進行文字情感分析, 2021年4月10日, 擷取自:

- <https://iter01.com/418571.html>
15. iT 邦幫忙 (2017), 處理影像的利器 -- 卷積神經網路, 2021 年 7 月 3 日, 擷取自:
<https://ithelp.ithome.com.tw/articles/10191820>
 16. iThome (2018), 機器學習: 特徵標準化, 2021 年 3 月 9 日, 擷取自:
<https://ithelp.ithome.com.tw/articles/10197357>
 17. Martin, S. (2018), CNN 與 RNN 之間的差異?, 2021 年 1 月 13 日, 擷取自:
<https://blogs.nvidia.com.tw/2018/09/05/whats-the-difference-between-a-cnn-and-a-n-rnn/>
 18. MBA 智庫百科 (2013), 股價分析, 2021 年 5 月 3 日, 擷取自:
<https://wiki.mbalib.com/zh-tw/%E8%82%A1%E4%BB%B7%E5%88%86%E6%9E%90>
 19. Medium (2017), [資料分析&機器學習] 第 1.2 講: Jupyter Notebook 介紹, 2021 年 2 月 20 日, 擷取自:
<https://medium.com/jameslearningnote/%E8%B3%87%E6%96%99%E5%88%86%E6%9E%90-%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-%E7%AC%AC1-2%E8%AC%9B-jupyter-notebook%E4%BB%8B%E7%B4%B9-705f023e3720>
 20. Mr.Marke 市場先生 (2020), 技術指標 是什麼? 常見的 技術指標 有哪些?, 2021 年 7 月 3 日, 擷取自:
<https://rich01.com/what-is-technical-indicator/>
 21. Python API 專欄 (2020), 何謂程式交易? 程式交易的優勢與如何學習, 2021 年 7 月 3 日, 擷取自:
<https://www.sinotrade.com.tw/richclub/PythonAPI/%E4%BD%95%E8%AC%82%E7%A8%8B%E5%BC%8F%E4%BA%A4%E6%98%93-%E7%A8%8B%E5%BC%8F%E4%BA%A4%E6%98%93%E7%9A%84%E5%84%AA%E5%8B%A2%E8%88%87%E5%A6%82%E4%BD%95%E5%AD%B8%E7%BF%92-5fdadffe3bbc823c4f4fff1>
 22. Sean Liu (2020), 量化交易 30 天 Day15 - 串接券商 API 做交易 (二) 台股歷史 tick 及分 K, 2020 年 12 月 20 日, 擷取自:
<https://ithelp.ithome.com.tw/m/articles/10244327>
 23. Summer (2018), 使用 Jupyter notebook 來寫 python, 2021 年 2 月 10 日, 擷取自:
<https://ithelp.ithome.com.tw/articles/10200046>
 24. YAHOO! 汽車 (無日期), 自動車, 2021 年 1 月 8 日, 擷取自:
<https://autos.yahoo.com.tw/>
 25. YAHOO! 汽車 (2020), Toyota 神級休旅車 RAV4 要出電動車啦! 純電動力版明年日本亮相, 2022 年進軍全球市場, 2021 年 07 月 05 日, 擷取自:
<https://autos.yahoo.com.tw/>
 26. Youngmi huang (2018), 使用 NLTK 搭配 Twitter API 拿取社群資料: 以川普的 Twitter 資料為例, 2021 年 1 月 10 日, 擷取自:
<https://reurl.cc/MZpXqX>
 27. 維基百科 (2021), Jupyter - 維基百科, 自由的百科全書 - Wikipedia, 2021 年 2 月 21 日, 擷取自:
<https://zh.wikipedia.org/wiki/Jupyter>
 28. 維基百科 (2021), 伊隆·馬斯克-維基百科, 自由的百科全書-Wikipedia, 2021 年 3 月 8 日, 擷取自:

- <https://zh.wikipedia.org/wiki/%E4%BC%8A%E9%9A%86%C2%B7%E9%A9%AC%E6%96%AF%E5%85%8B>
29. 維基百科 (2021), 深度學習, 2021 年 7 月 3 日, 擷取自:
<https://zh.wikipedia.org/wiki/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0>
 30. 維基百科 (無日期), 長短期記憶, 2021 年 3 月 23 日, 擷取自:
<https://zh.wikipedia.org/wiki/%E9%95%B7%E7%9F%AD%E6%9C%9F%E8%A8%98%E6%86%B6>
 31. 維基百科 (無日期), 比特幣, 2021 年 7 月 3 日, 擷取自:
<https://zh.wikipedia.org/wiki/%E6%AF%94%E7%89%B9%E5%B8%81>
 32. 財經 PYTHON 教學 (2020), Python 新手教學 (Part 1): 用爬蟲爬全球股價!, 2021 年 1 月 10 日, 擷取自: <https://reurl.cc/MZpXan>
 33. 鉅亨網新聞中心 (2020), IDC: 中國新能源車今年銷量約 116 萬輛 未來 5 年飛速成長, 2021 年 1 月 5 日, 擷取自: <https://news.cnyes.com/news/id/4547076>
 34. 鉅亨網新聞中心 (2020), IDC: 中國新能源車今年銷量約 116 萬輛 未來 5 年飛速成長, 2021 年 1 月 5 日, 擷取自: <https://news.cnyes.com/news/id/4547076>
 35. 鉅亨網新聞中心 (2021), 特斯拉財報前瞻: 交車 50 萬輛無懸念、FSD 訂閱制可望推出!, 2021 年 1 月 8 日, 擷取自:
<https://news.cnyes.com/news/id/4563389>
 36. 科技新報 (2020), 2020 年特斯拉無意外稱霸純電車市場, 插電混動車市場歐系大亂鬥, 2021 年 1 月 10 日, 擷取自:
https://technews.tw/2020/12/15/tesla-lead-bev-market-phev-4brand-compete/?fbclid=IwAR1xsemF_yzmXBo0kLzaIzvFfgS3UWVcpOcoo_0uHtkK6-OC9o5BkvWP_rdk
 37. 天下雜誌 (2020), 圖解特斯拉供應鏈 | 29 家代表台廠出列! 3 優勢讓馬斯克離不開台灣, 2021 年 1 月 10 日, 擷取自:
https://www.cw.com.tw/article/5101653?template=transformers&fbclid=IwAR1ODZBL6ilEt9tEe-rWLIZUH6iKuB2zfqMHLLeVstGiL_pCszlZ5kD645_c
 38. 皮旭庭 (2018), 深度學習及其在凝態物理上的應用 - 中篇: CNN 與 RNN, 2021 年 7 月 3 日, 擷取自:
https://pb.ps-taiwan.org/catalog/ins.php?index_m1_id=5&index_id=402
 39. 量化研究院 (2019), [Python 股票爬蟲]十分鐘學會 Python 爬台股, 2021 年 3 月 8 日, 擷取自: <https://ushluap.github.io/2019-06-26-crawler-1/>
 40. 星展銀行 (2020), 每日金融及外匯市場晨訊, 2021 年 1 月 12 日, 擷取自:
https://www.dbs.com.tw/iwov-resources/pdf/legal%20disclaimers%20and%20announcements/01_personal%20finance/06_asian%20insights/FX20200109.htm
 41. 台灣證卷交易所 (無日期), 上市股價, 2021 年 1 月 20 日, 擷取自:
<https://www.twse.com.tw/zh/>
 42. 程式前沿 (2019), Jupyter Notebook 介紹與入門| 程式前沿, 2021 年 3 月 8 日, 擷取自:
<https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/451444/>
 43. 中國乘聯會 (2021), IDC: 未來五年中國新能源汽車市場將強勁增長, 2021 年 1 月 5 日, 擷取自:
<https://new.qq.com/omn/20201204/20201204A03SFG00.html>

44. 數位時代 (2020), Model 3 蟬聯電動車銷量冠軍! 這輛 20 萬有找的小車成大黑馬, 魅力在哪裡?, 2021 年 7 月 3 日, 擷取自:
<https://www.bnext.com.tw/article/60394/tesla-popularity-model-3-model-y-october-2020>
45. 數位時代 (2021), 馬斯克估重新接受比特幣, 要求礦工使用 50% 乾淨能源! 比特幣漲回 4 字頭, 2021 年 7 月 6 日, 擷取自:
<https://www.bnext.com.tw/article/63364/elon-musk-may-accept-bitcoin-again>
46. 數位時代 (2021), 特斯拉暫停比特幣支付購車, 馬斯克: 尋找能源耗損率不到 1% 的密碼貨幣, 2021 年 7 月 6 日, 擷取自:
<https://www.bnext.com.tw/article/62832/elon-musk-says-tesla-will-stop-accepting-bitcoin-for-car-purchases>
47. Mr. Marke 市場先生 (2020), 股價淨值比 (PB 比) 是什麼? 公式如何計算、和本益比有什麼差別?, 2021 年 7 月 21 日, 擷取自:
<https://rich01.com/what-is-pb-ratio/>
48. 慢活夫妻 George&Dewi (2021), 【股息殖利率越高越好?】小心中了高股息的「這個」陷阱!, 2021 年 7 月 21 日, 擷取自:
<https://george-dewi.com/high-dividend-yield/#t-1613014317100>
49. 周商 (2021), 特斯拉已砸 15 億美元買比特幣、或將長期持有數位資產, 2022 年 2 月 20 日, 擷取自:
<https://www.businessweekly.com.tw/business/blog/3005519>
50. 華盛頓郵報 (2021), 馬斯克口無遮攔, 特斯拉「熊出沒」!, 2022 年 2 月 22 日, 擷取自:
<https://tw.news.yahoo.com/news/%E9%A6%AC%E6%96%AF%E5%85%8B%E5%8F%A3%E7%84%A1%E9%81%AE%E6%94%94-%E7%89%B9%E6%96%AF%E6%8B%89-%E7%86%8A%E5%87%BA%E6%B2%92-220429265.html>
51. BBC (2021), 推特民意調查支持馬斯克出售特斯拉 10% 股份, 2022 年 2 月 22 日, 擷取自:
<https://www.bbc.com/zhongwen/trad/world-59203825>
52. 鉅亨 (2020), 股價太高了, 特斯拉應聲暴跌 10%, 2022 年 2 月 23 日, 擷取自:
<https://news.cnyes.com/news/id/4470181>
53. 數位時代 (2020), 一篇貼文讓特斯拉市值蒸發百億美元, 馬斯克又踩到監管紅線?, 2022 年 2 月 25 日, 擷取自:
<https://www.bnext.com.tw/article/57541/elon-musk-tesla-twitter>

二、英文部分

1. Basil K Jose. Twint: Twitter Scraping Without Twitter's API. (2021). Retrieved 18 Feb. 2021, from
<https://medium.com/analytics-vidhya/twint-twitter-scraping-without-twitters-api-aca8ba1b210e>
2. Bercovici, J. Twitter Chatter Can Predict Stock Prices, Study Finds. (2012). Retrieved 18 Feb. 2021, from
<https://www.forbes.com/sites/jeffbercovici/2012/03/22/twitter-chatter-can-predict-stock-prices-study-finds/?sh=36436dc44070>
3. CoinDesk. Bitcoin. (no date). Retrieved 6 July. 2021, from
<https://www.coindesk.com/price/bitcoin>
4. Colah's blog. Understanding LSTM Networks. (2015). Retrieved 17 March. 2021,

- from <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
6. InsideEVs. InsideEVs. (no date). Retrieved 8 Jan. 2021, from <https://insideevs.com/>
 7. Investing.com. Tesla Inc (TSLA). (no date). Retrieved 3 July 2021, from <https://hk.investing.com/equities/tesla-motors-historical-data>
 8. Sproul Himself. GitHub Tesla Stock Analysis from Elon Musk's Tweets Using NLP. (2019). Retrieved 5 Jan. 2021, from https://github.com/SproulHimself/TSLA_stock_twitter_sentiment_analysis
 9. Tal Rodin. GitHub Elon-Musk-Tweets-Visualization. (2019). Retrieved 5 Jan. 2021, from <https://github.com/TalRodin/Elon-Musk-Tweets-visualization>





附錄一 畢業專題 系統操作手冊

【專題執行計畫表】

組名	馬斯克推文之情感分析		
組員	班級	學號	姓名
	資四 A	10710147	張雅喬
	資四 A	10710149	林以晴
	資四 A	10710108	陳婕名
	資四 A	10710115	黃乙茜
	資四 A	10710120	施鈺函
	資四 A	10710126	謝滄瑾
專題名稱	馬斯克推文之情感分析		
專題資訊系統功能描述			
<p>本專題主要以分析馬斯克的網路言論是否對於特斯拉之後 24、48、72 小時的股價有所影響，全部用自然語言處理、情感分析、LSTM 深度學習，將資料分為訓練、驗證、測驗模型得知準確度。從這些蒐集的數據中，拿取部分數據來當作測試資料，以測試模型的結果。而當測試模型結果跟驗證模型數據符合時，方能取得最佳狀態之模型，在這樣重複性的交叉驗證下，增強了本專題研究的精準度。</p>			
指導老師簽名	王嫻惠 副教授	日期	2020 年 10 月 13 日
備註			

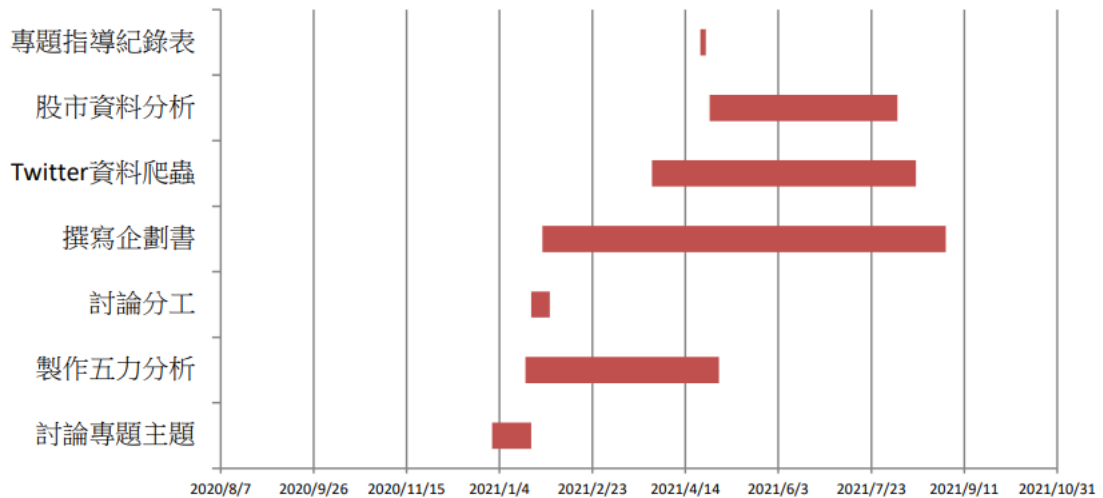
【專題工作分配表】

組名	馬斯克推文之情感分析		填寫人	陳婕名		
專題名稱	馬斯克推文之情感分析		填寫日期	2021年10月10日		
	張雅喬	林以晴	陳婕名	黃乙茜	謝涓瑾	施鈺函
查找收集 相關資料	v	v	v	v	v	v
工作分配 意見整合	v	v				
確立目標 內容探討	v	v	v	v	v	v
程式設計	v	v		v	v	
股市分析					v	v
企劃書			v			v
會議記錄			v			
PPT 製作				v	v	
專題報告者	v	v				

【GANTT 圖】

組名	馬斯克推文之情感分析	填寫人	陳婕名
專題名稱	馬斯克推文之情感分析	填寫日期	2020年5月26日

任務名稱	開始時間	完成時間	任務工期
討論專題主題	2020/12/31	2021/1/20	21
製作五力分析	2021/1/18	2021/5/2	104
討論分工	2021/1/21	2021/1/30	10
撰寫企劃書	2021/1/27	2021/8/31	217
Twitter資料爬蟲	2021/3/27	2021/8/15	142
股市資料分析	2021/4/27	2021/8/5	101
專題指導紀錄表	2021/4/22	2021/4/24	3



【開發工具清單】

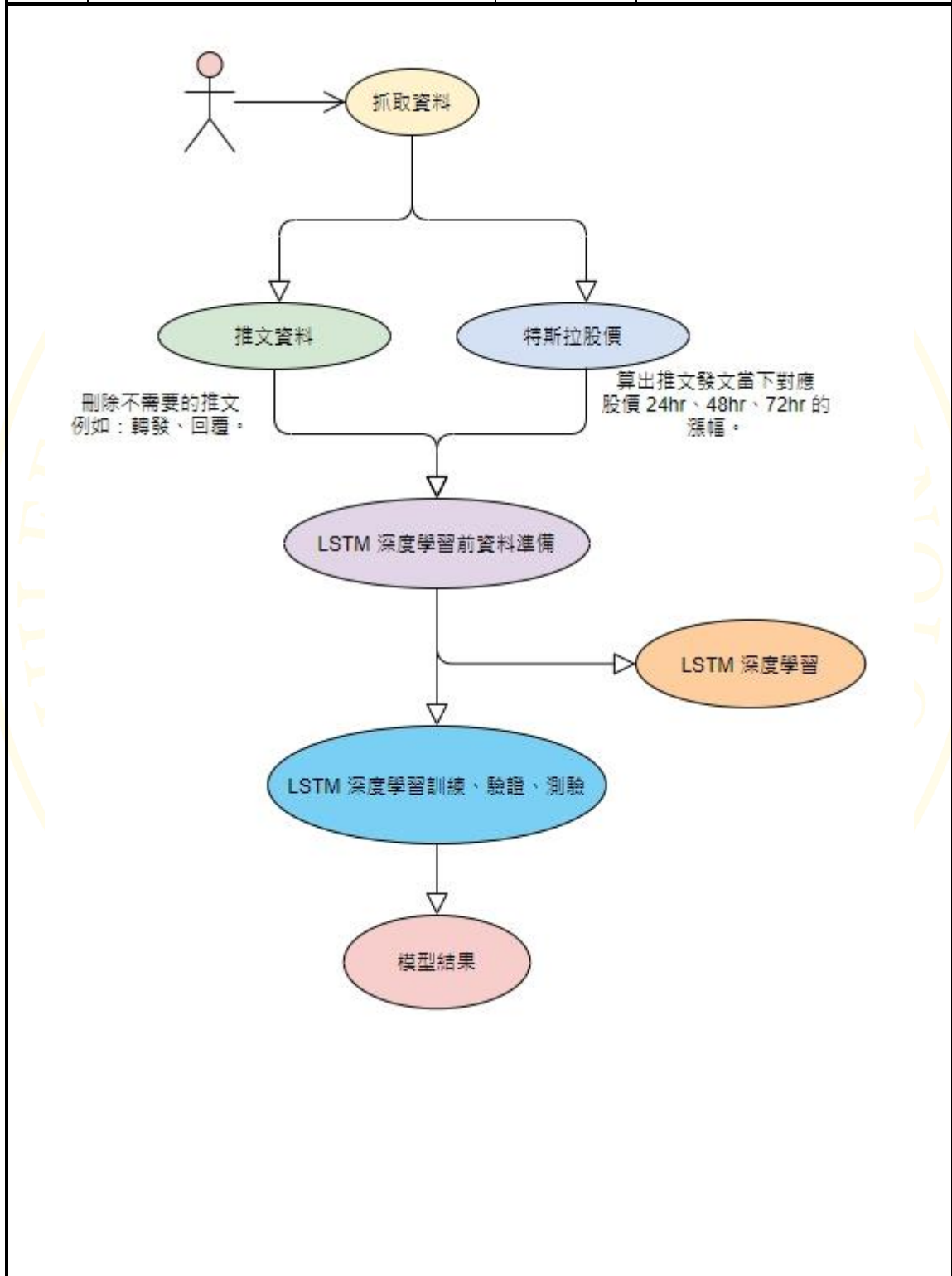
組名	馬斯克推文之情感分析	填寫人	施鈺函
專題名稱	馬斯克推文之情感分析	填寫日期	2021年06月13日
名稱		功能	
jupyter notebook		編寫程式分析、股票分析	
TWINT		用來爬蟲	



The image contains a large, faint watermark of the Chihlee University of Technology logo. The logo is circular with the text 'CHIHLEE UNIVERSITY OF TECHNOLOGY' around the top and 'SINCE 1965' at the bottom. In the center is a shield-shaped emblem divided into three sections: a blue triangle on the left with the Chinese character '理' (Li), a red triangle on the right with the Chinese character '致' (Zhi), and a white triangle at the bottom with the Chinese character '大' (Da). The character '學' (Xue) is positioned below the shield, and the characters '理致大學' (Li Zhi University) are arranged vertically in the center.

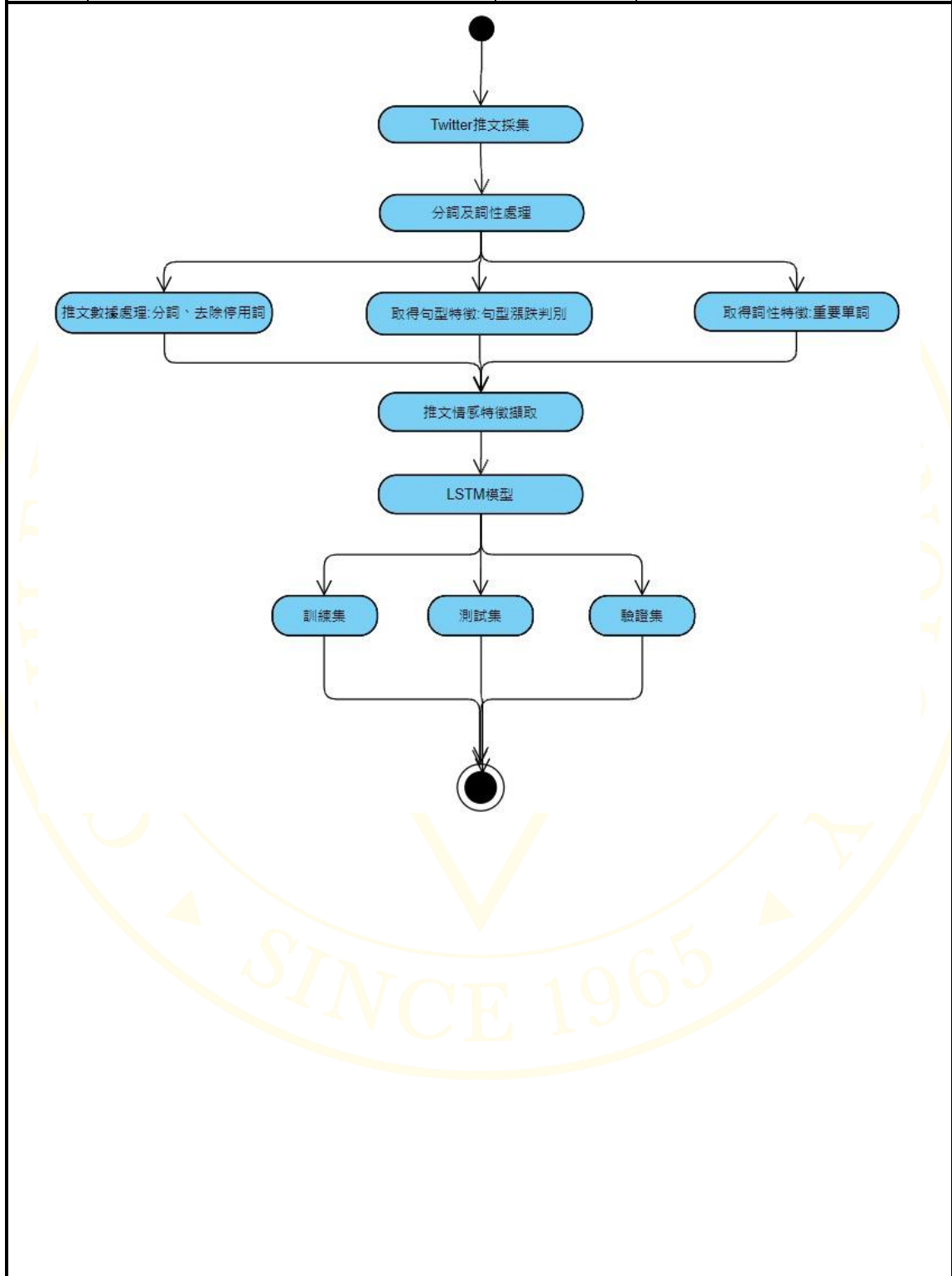
【使用個案圖】

組名	馬斯克推文之情感分析	填寫人	施鈺函
專題名稱	馬斯克推文之情感分析	填寫日期	2021年10月02日



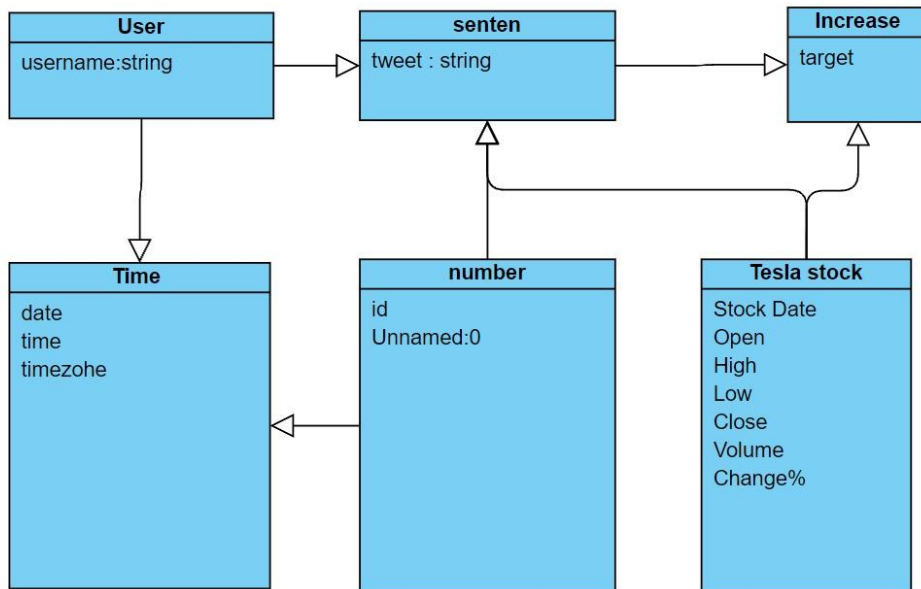
【活動圖】

組名	馬斯克推文之情感分析	填寫人	陳婕名
專題名稱	馬斯克推文之情感分析	填寫日期	2021年10月03日



【類別圖】

組名	馬斯克推文之情感分析	填寫人	陳婕名
專題名稱	馬斯克推文之情感分析	填寫日期	2021年10月03日



【使用者操作手冊】

組名	馬斯克推文之情感分析	填寫人	施鈺函
專題名稱	馬斯克推文之情感分析	填寫日期	2021年06月24日
畫面編號	01	畫面名稱	主畫面
系統畫面	 <p>系統不適用</p>		
操作說明	<p>系統不適用</p>		

【測試相關計畫】

組名	馬斯克推文之情感分析	填寫人	施鈺函
專題名稱	馬斯克推文之情感分析	填寫日期	2021 年 06 月 24 日

1. Analyze data 分析資料
利用馬斯克的推文和 Tesls 的股價分析，透過情感分析來判斷股市在各個時間點上是否有漲跌。

2. Design test strategy 設計測試策略
在測試過程中要達成的目標是能夠情感分析股市，並且達到精準度 0.XX，但遇到準確度不高的問題，我們會繼續修改程式、練習模型、標記出有問題的部分。

3. Define test objectives 定義測試目標
測試的內容有：畫面的呈現、資料的輸入、推文及時間的資料、預測的成效。

4. Define test criteria 定義測試規範
我們制定了測試的目標，例如：精準度達 0.92、情感分析的正反面向、股市的漲跌幅，經過不斷的測試、修改，直到通過目標為止。

5. Resource planning 資源計畫
執行程式所需要的 Python 資源包含爬蟲、機器學習及 NLP、自然語言處理工具、數據分析工具。

6. Schedule & Estimation 時程與估計
研究此情感分析需要在 5 個月內做完 Twitter 資料爬蟲、4 個月內抓取股市資料分析，以及在剩下 1 個月的時間經過不斷的訓練模型、修改 Bug。

7. Determine test deliverables 決定測試交付
在測試前需準備股市資料、Jupyter 分析環境，以便做測試準備。在測試中期要檢測是否能分析股市及精準度，再透過訓練模型，達成準確度的提升。

【專案結案報告】

組名	馬斯克推文之情感分析	填寫人	施鈺函
專題名稱	馬斯克推文之情感分析	填寫日期	2021年06月24日
<p>1.首先要感謝專題指導老師王嫵惠副教授以及班導師曹祥雲教授，經過老師的細心指導及建議下，讓我們的專題能夠完成。</p> <p>2.在情感分析的部分，我們達到了以程式實際分析的目標。</p> <p>3.在實際執行中，原本預計預測值能達成 0.92，但是在程式撰寫的過程中，預測值停滯不前，經過不斷的更改、訓練部分的程式，且不斷的訓練模型，讓預測值達成比預計的數值來的還要高。</p> <p>4.在決定目標時，我們都沒有聽過何謂情感分析的方法，在導師的帶領下讓我們一步一步的了解、思考、執行，呈現我們的目標在專題上，也代表著我們只要有心就能突破我們所不知道的領域，也能做出成品來展示給大家看。</p> <p>5. 在情感分析上對於主觀和客觀文字的分析並不是很精準，以及 LSTM 要解決 RNN 所產生出短期記憶的問題，希望未來的發展方向能夠前往長期記憶以及更加精準的目標前進。</p>			

【會議記錄】

專題名稱	馬斯克推文之情感分析				
會議編號	01	召集人兼主席	張雅喬	紀錄者	陳婕名
討論主題	首次開會決定專題方向			會議時間	2021/01/21
				會議地點	致理科技大學
上次會議	決議事項		執行狀況		
	首次開會		首次開會		
本次會議	本週工作進度		本週工作內容		負責人員
	1. 選定主題方向 2. 程式抓取資料方向		1. 決定要以哪位具有影響力（股價 or 比特幣）的人物為分析對象 2. 分配工作項目 3. 利用 python 來抓取 Twitter 的文章 4. 分析抓取的文章會影響到什麼社會、經濟波動 程式撰寫 1. 資料爬蟲（Twitter） 2. 寫進資料庫 3. 抓取什麼的標地對象（股價 or 比特幣）		林以晴 陳婕名 黃乙茜 施鈺函 謝涓瑾 張雅喬
本次會議內容	這次開會目的為設立主題目標為主，老師請我們尋找什麼樣的人物足以影響一個目標（股價 or 比特幣）比如：川普，經過討論後，決定了特斯拉創辦人馬斯克作為分析對象，並五力分析特斯拉電動車的優、劣，其中包括了「來自買方的議價能力、來自供應商的議價能力、來自潛在進入者的威脅、來自替代品的威脅以及來自現有競爭者的威脅」並調查現在誰是有出電動車的汽車廠牌。				
決議事項（與主席裁示）					
請各位利用這的時間來決定我們要分析的目標，並找出目標人物及股票影響的資訊，給予參考利用。					
請簽名	請簽名	請簽名	請簽名	請簽名	請簽名
下次會議	召集人	張雅喬	紀錄者	陳婕名	時間
					2021/01/21
					地點
					導師辦公室
預定討論主題	報告老師目前選定的追蹤對象，以及他的文章會影響什麼。				
指導老師意見	盡快完成 Twitter 資料的抓取，才能執行後續動作。				

專題名稱	馬斯克推文之情感分析				
會議編號	02	召集人兼主席	張雅喬	紀錄者	陳婕名
討論主題	確立主題方向			會議時間	2021/1/29
				會議地點	致理科技大學
上次會議	決議事項	執行狀況			
	1.選定主題方向 2.程式抓取資料方向	1.決定要以哪位具有影響力(股價 or 比特幣)的人物為分析對象 2.分配工作項目 3.分析抓取的文章會影響到什麼社會、經濟波動 程式撰寫 1.資料爬蟲(Twitter) 2.寫進資料庫 3.抓取什麼的標地對象(股價 or 比特幣)			
本次會議	本週工作進度	本週工作內容		負責人員	
	1.抓 Twitter 推文資料 2.蒐集某家股價波動 Ex: 特斯拉、比特幣 3.分工名單	1.蒐集馬斯克過去在 Twitter 的推文 2.特斯拉以前的股價(每分鐘),分析回測以前說的話影響了什麼 3.NLTK 抓取馬斯特所有資料 4.五力分析廠商		林以晴 陳婕名 黃乙茜 施鈺函 謝涓瑾 張雅喬	
本次會議內容	目前找出了有在製作目前有做純電車的廠商: 特斯拉、福斯、Nissan 以及插電混合式電動車: BMW、Volvo、賓士、Toyota 這幾家廠商, 透過這些廠商跟特斯拉做對比, 進而完成五力分析, 五力分析包括 1.來自買方的議價能力 2.來自供應商的議價能力 3.來自潛在進入者的威脅 4.來自替代品的威脅 5.來自現有競爭者的威脅, 並分析股價影響。				
決議事項 (與主席裁示)					
請各位這週盡快把分工名單交出來, 以及找出馬斯克言論情感分析對五力分析之影響。					
請簽名	請簽名	請簽名	請簽名	請簽名	請簽名
下次會議	召集人	張雅喬	紀錄者	陳婕名	時間 2021/02/05
					地點 導師辦公室
預定討論主題	進行開發環境建置				
指導老師意見	目前先完成五力分析資料蒐集, 及特斯拉歷史股價抓取。				

【會議記錄】

專題名稱	馬斯克推文之情感分析				
會議編號	04	召集人兼主席	張雅喬	紀錄者	陳婕名
討論主題	進行開發環境建置			會議時間	2021/2/05
				會議地點	致理科技大學
上次會議	決議事項		執行狀況		
	1. 抓 Twitter 推文資料 2. 蒐集某家股價波動 Ex：特斯拉、比特幣 3. 分工名單		1. 蒐集馬斯克過去在 Twitter 的推文 2. 特斯拉以前的股價（每分鐘），分析回測以前說的話影響了什麼 3. NLTK 抓取馬斯特所有資料 4. 五力分析廠商		
本次會議	本週工作進度		本週工作內容		負責人員
	1. 企劃書、會議紀錄 2. 抓 Twitter 推文資料 3. 更新股市資料		1. 企劃書前置作業 2. 一篇會議紀錄 3. 繼續抓取 Twitter 推文資料 4. 抓取其他競爭業者的股市數據、或比特幣		林以晴 陳婕名 黃乙茜 施鈺函 謝涓瑾 張雅喬
本次會議內容	企劃書的部份要開始用了，不久之後就要繳交企劃書，在交之前必須先給指導老師看過一遍，再來就是要開始找比賽活動參加。				
決議事項（與主席裁示）					
目前在進度的部分可能要加快速度，希望在股市資料部分時間上能夠抓取的精確一點。					
請簽名	請簽名	請簽名	請簽名	請簽名	請簽名
下次會議	召集人	張雅喬	紀錄者	陳婕名	時間
					2021/02/25
				地點	導師辦公室
預定討論主題	目前程式建置中遇到的 bug 及問題排解				
指導老師意見	要驗證什麼訊息對甚麼股價偏正向或負向，所以幾個負責找資料要互相配合，確認其中時間欄位是對相同時區。				

【會議記錄】

專題名稱	馬斯克推文之情感分析					
會議編號	05	召集人兼主席	張雅喬	紀錄者	陳婕名	
討論主題	目前程式建置中遇到的 bug 及問題排解			會議時間	2021/02/25	
				會議地點	導師辦公室	
上次會議	決議事項		執行狀況			
	1.企劃書、會議紀錄 2. 抓 Twitter 推文資料 3.更新股市資料		1.企劃書前置作業 2. 一篇會議紀錄 3.繼續抓取 Twitter 推文資料 4.抓取其他競爭業者的股市數據、或比特幣			
本次會議	本週工作進度		本週工作內容		負責人員	
	1.企劃書、會議紀錄 2. 更新並整理推文資料成 CSV 檔 3.更新股市資料到 2015 年以後		1.撰寫企劃書、會議紀錄 2.抓取更早的 Twitter 推文資料 3.開始測試 nltk 情感分析功能 4.股價抓取每日漲跌、開盤收盤價		林以晴 陳婕名 黃乙茜 施鈺函 謝洄瑾 張雅喬	
本次會議內容	Twitter 目前抓到 2021/01/02 的推文資料，可以再試試往前抓到 2015 年，配合股價資料。如果只有 2021/01/02 到現在，股價資料只有約 100 筆，會太少，建議至少有 1000 筆資料。另外，Twitter 這邊也要開始試試 nltk 的情感分析功能，股價也要開始處理，計算每天的漲跌幅、開盤價、收盤價、上下影線長度，準備當學習模型的目標。					
決議事項（與主席裁示）						
程式組開始處理 nltk 的情感分析功能，股價組開始抓取股價開盤、收盤、最高、最低跟交易量，對照資料互相配合。						
請簽名	請簽名	請簽名	請簽名	請簽名	請簽名	
下次會議	召集人	張雅喬	紀錄者	陳婕名	時間	2021/03/08
					地點	導師辦公室
預定討論主題	開始撰寫情感分析的程式部分，預測股價資料整理成 CSV 檔輸出					
指導老師意見						

專題名稱	馬斯克推文之情感分析				
會議編號	06	召集人兼主席	張雅喬	紀錄者	陳婕名
討論主題	收尾階段			會議時間	2021/03/08
				會議地點	導師辦公室
上次會議	決議事項		執行狀況		
	1.企劃書、會議紀錄 2.更新並整理推文資料成 CSV 檔 3.更新股市資料到 2015 年以後		1.撰寫企劃書、會議記錄 2.抓取更早的 Twitter 推文資料 3.開始測試 nltk 情感分析功能 4.股價抓取每日漲跌、開盤收盤價		
本次會議	本週工作進度		本週工作內容		負責人員
	1.企劃書修正 2.製作專題發表 ppt 3.專題海報設計		1.企劃書修正 2.製作專題發表 ppt 3.專題海報設計 4.所有書面內容修正		林以晴 陳婕名 黃乙茜 施鈺函 謝洵瑾 張雅喬
本次會議內容	目前所有工作都到一個段落了，請各位組員把手上的工作確認一下是否都有完成，組員之間也互相幫忙檢查是否有誤，最後再給老師過目。				
決議事項（與主席裁示）					
再繳交報告截止前完成所有的資料並繳交					
請簽名	請簽名	請簽名	請簽名	請簽名	請簽名
下次會議	召集人	張雅喬	紀錄者	陳婕名	時間
					2021/04/20
				地點	導師辦公室
預定討論主題	開始準備專題發表會				
指導老師意見					

符號（公式）說明

Symbol	Meaning
Θ	Debye's constant or characteristic temperature
Ω	efficiency; number of molecules
Ψ	availability of a closed system
Δ	internal energy (change) of reaction
Φ	availability of a closed system
I	specific irreversibility
λ	critical state
M	Joule-Thomson coefficient
N	stoichiometric coefficient (number of moles in chemical equation)
E	cutoff ratio

