

國立臺灣師範大學
資訊工程研究所碩士論文

指導教授：吳榮根 博士
共同指導教授：葉耀明 博士

應用開放資料預測農產品菜價之研究：



Applying Open Data to Predict Vegetable Prices in
Farm Products: Taking Cabbage as an Example

研究生：翟柏森 撰

中 華 民 國 一 零 七 年 八 月

摘要

農產品價格的波動影響著我們日常生活。近年來劇烈的氣候變化更加劇菜價漲跌。本研究目的為農產品交易資訊透明化做出貢獻，研究流程分成三大部分：自動化資料擷取與特徵工程、資料視覺化、農產品價格預測模型。本研究首先搜集農產品交易開放資料、氣象開放資料和颱風警報資料，並透過資料清洗及各項特徵工程方法來整理資料和特徵建構，經過多次試算與調整後，再將資料擷取、特徵工程等步驟整合並撰寫成自動化擷取程式，使得未來更新資料能夠不依靠人工便可自動化更新；此外本論文提出農產品交易訊息視覺化方法，藉由視覺化圖形彼此交互比對，使得大眾能夠直觀地觀察、分析龐大且繁雜的數據，最後使用類神經網路之 LSTM（長短期記憶模型）設計價格預測模型，預測全國最大宗蔬菜-甘藍價格。

關鍵字：政府開放資料、農產品、蔬菜價格預測、類神經網路

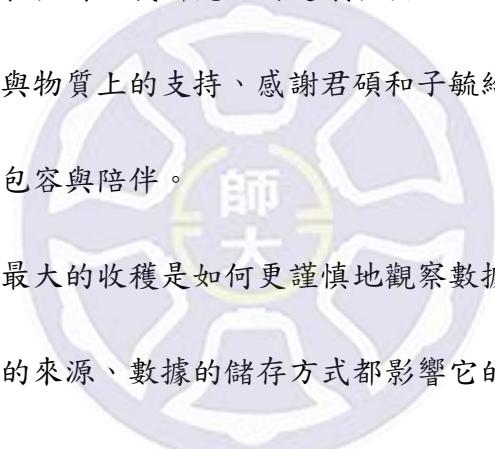
Abstract

The purpose of this study is to predict vegetable prices in the wholesale markets using governmental open data. We collected various governmental open data in Taiwan, including agricultural product prices in wholesale markets and climate statistics of each county in Taiwan. In order to develop our prediction model, we organize and construct vegetable price data features using data cleansing and feature engineering. We also develop automated data extraction programs to constantly download the new data without human intervention. In addition, a visualization scheme is developed for these complicated agricultural transaction data. Our prediction model, LSTM (Long short-term memory) architecture, is developed using Deep Learning RNN (Recurrent Neural Network) architecture. A preliminary vegetable price prediction experiment is conducted.

Index Terms— government open data, vegetable price prediction, deep learning

誌謝

轉眼研究所即將畢業，研究這段路是漫長、孤單而充實的，過程中最開心的是呈現出一些半成品、實驗有好的結果，可以發現一些資料背後的故事。完成碩士論文首先要先感謝 葉耀明教授，葉老師無論是在研究實驗、論述撰寫、邏輯思考上都給予我專業的協助與指導，研究之餘還會分享許多生活處事態度的經驗，非常感謝教授對於我的各種幫助與教導，收穫良多、一言難盡；此外感謝口試教授對於我研究上的建議和指正；也感謝家人在我研究所求學階段給予我精神與物質上的支持、感謝君碩和子毓給予我研究上的建議和鼓勵，還有生活上的包容與陪伴。



完成這份論文最大的收穫是如何更謹慎地觀察數據和資料，因為數據並非絕對客觀的：數據的來源、數據的儲存方式都影響它的真實性；研究的痛苦在於知識太少、做太多、做太少；更痛苦的是太多想法來不及記下來、記下來會永遠做不完。那讓我從懷疑一切，到放下一切選擇先相信。

最後期望未來有更多人能一起關心農產品相關議題，一起支持在台灣各地、各領域辛勤耕耘的每個角色。

翟柏森 107,7,26

汀州路四段 85 巷

目錄

| | |
|-----------------------|------|
| 摘要 | i |
| Abstract | ii |
| 誌謝 | iii |
| 目錄 | iv |
| 附表目錄 | viii |
| 附圖目錄 | ix |
| 第一章、緒論 | 1 |
| 1.1 研究背景 | 1 |
| 1.2 研究目的 | 2 |
| 1.3 研究架構 | 4 |
| 第二章、文獻探討 | 6 |
| 2.1 開放資料 | 6 |
| 2.2 特徵工程 | 6 |
| 2.3 隨機森林演算法 | 7 |
| 2.4 類神經網路之 LSTM | 7 |
| 2.5 植物生長理論 | 8 |
| 第三章、資料擷取與特徵工程 | 10 |
| 3.1 開放資料擷取 | 10 |

| | |
|---------------------------------|-----------|
| 3.2 農產品交易開放資料之特徵工程 | 11 |
| 3.2.1 建立新欄位 | 12 |
| 3.2.2 刪除市場休息資料 | 14 |
| 3.2.3 轉換欄位名稱為英文 | 15 |
| 3.2.4 刪除相依欄位 | 16 |
| 3.2.5 建構日期特徵與市場休息特徵 | 17 |
| 3.3 氣象開放資料之特徵工程..... | 19 |
| 3.3.1 特徵工程 | 22 |
| 3.3.2 獨立各縣市天氣特徵 | 23 |
| 3.4 建立自動化資料擷取程式..... | 26 |
| 第四章、資料視覺化..... | 27 |
| 4.1 農產品供貨量分析 | 28 |
| 4.1.1 農產品供貨量前 30 名分析 | 29 |
| 4.1.2 蔬菜供貨量前 30 名分析 | 31 |
| 4.2 市場供貨量分析 | 33 |
| 4.2.1 全部市場供貨量分析 | 34 |
| 4.2.2 蔬菜市場分析 | 35 |
| 4.3 單一品項交易訊息分析 | 36 |
| 4.3.1 各年總供貨量比例 | 36 |

| | |
|-----------------------------|-----------|
| 4.3.2 歷年供貨量及平均價 | 36 |
| 4.3.3 每日交易訊息分析 | 38 |
| 4.4 五大市場分析 | 39 |
| 4.5 多品項交易訊息分析 | 40 |
| 4.6 關聯度分析 | 41 |
| 4.6.1 國產蔬菜供貨量關聯度分析 | 41 |
| 4.6.2 國產蔬菜平均價關聯度分析 | 43 |
| 4.6.3 進階多品項分析 | 44 |
| 4.7 供貨量與天氣特徵分析 | 48 |
| 第五章、特徵分析..... | 52 |
| 5.1 平均價漲跌分類方法 | 53 |
| 5.2 供貨量特徵 | 55 |
| 5.3 天氣特徵 | 56 |
| 5.3.1 溫度特徵 | 56 |
| 5.3.2 降水量特徵 | 61 |
| 5.4 颱風警報發布特徵 | 66 |
| 5.5 市場休息與日期特徵 | 67 |
| 第六章、價格預測模型 | 69 |
| 6.1 預測未來 5 日價格模型 | 69 |

| | |
|----------------------|----|
| 6.2 價格預測混合模型 | 71 |
| 6.3 價格預測模型結果討論 | 76 |
| 6.4 研究結果與未來展望 | 78 |
| 附錄 | 80 |
| 參考文獻..... | 89 |



附表目錄

| | |
|----------------------------|----|
| 表 3.1 開放資料來源..... | 10 |
| 表 3.2 農產品交易資料集訊息..... | 11 |
| 表 3.3 農產品資料-欄位中英文對照表 | 15 |
| 表 3.4 市場休息特徵變換結果..... | 19 |
| 表 3.5 氣象測站代碼欄位說明..... | 20 |
| 表 3.6 氣象開放資料集訊息..... | 21 |
| 表 3.7 氣象資料-欄位中英文對照表 | 22 |
| 表 5.1 價格漲跌分類規則..... | 54 |
| 表 5.2 累積供貨量實驗結果..... | 55 |
| 表 5.3 當日溫度實驗結果..... | 56 |
| 表 5.4 累積 30 日溫度實驗結果..... | 57 |
| 表 5.5 累積 60 日溫度實驗結果..... | 59 |
| 表 5.6 當日降水量實驗結果..... | 61 |
| 表 5.7 累積 30 日降水量實驗結果..... | 62 |
| 表 5.8 累積 60 日降水量實驗結果..... | 64 |
| 表 5.9 颱風警報特徵實驗結果..... | 66 |
| 表 5.10 休市與日期特徵實驗結果..... | 68 |

附圖目錄

| | |
|-------------------------------|----|
| 圖 1.1 研究架構圖..... | 5 |
| 圖 3.1 農產品市場代碼表..... | 12 |
| 圖 3.2 農產品開放資料集..... | 13 |
| 圖 3.3 農產品資料-建立新欄位結果 | 14 |
| 圖 3.4 農產品資料-市場休息資料 | 15 |
| 圖 3.5 農產品資料-最終格式 | 16 |
| 圖 3.6 農產品資料-整理前訊息 | 16 |
| 圖 3.7 農產品資料-整理後訊息 | 17 |
| 圖 3.8 日期資料集..... | 18 |
| 圖 3.9 合併農產品交易與日期資料集結果..... | 18 |
| 圖 3.10 休市特徵資料集..... | 19 |
| 圖 3.11 氣象測站代碼資料集..... | 21 |
| 圖 3.12 氣象資料集..... | 22 |
| 圖 3.13 氣象資料-整理後訊息 | 23 |
| 圖 3.14 各縣市天氣特徵-溫度 | 24 |
| 圖 3.15 溫度折線圖-全部縣市 | 25 |
| 圖 3.16 溫度折線圖-彰化縣、台中市 | 25 |
| 圖 4.1 資料視覺化流程圖 | 27 |
| 圖 4.2 農產品供貨量前 30 名統計..... | 29 |
| 圖 4.3 農產品供貨量前 30 名圓餅圖..... | 30 |
| 圖 4.4 農產品供貨量前 30 名長條圖..... | 30 |
| 圖 4.5 胡蘿蔔-清洗 SB2 供貨量長條圖 | 31 |
| 圖 4.6 蔬菜供貨量前 30 名圓餅圖..... | 32 |

| | |
|---|----|
| 圖 4.7 蔬菜供貨量前 30 名長條圖..... | 32 |
| 圖 4.8 甘藍-改良種 LA2 供貨量長條圖 1..... | 33 |
| 圖 4.9 甘藍-改良種 LA2 供貨量長條圖 2..... | 33 |
| 圖 4.10 全部市場供貨量圓餅圖..... | 34 |
| 圖 4.11 全部市場供貨量長條圖..... | 34 |
| 圖 4.12 蔬菜市場供貨量圓餅圖..... | 35 |
| 圖 4.13 蔬菜市場供貨量長條圖..... | 35 |
| 圖 4.14 甘藍-初秋 LA1 歷年總供貨量圓餅圖 | 36 |
| 圖 4.15 甘藍-初秋 LA1 歷年供貨量（堆積圖）及平均價 | 37 |
| 圖 4.16 甘藍-初秋 LA1 歷年供貨量（平坦圖） | 37 |
| 圖 4.17 甘藍-初秋 LA1 每日交易訊息分析 | 38 |
| 圖 4.18 甘藍-初秋 LA1 每日交易訊息分析-隱藏休市訊息..... | 38 |
| 圖 4.19 甘藍-初秋 LA1 五大市場分析 | 39 |
| 圖 4.20 胡蘿蔔之多品項交易訊息分析..... | 40 |
| 圖 4.21 國產蔬菜供貨量關聯度分析..... | 42 |
| 圖 4.22 國產蔬菜供貨量關聯度分析-篩選低關聯度 | 42 |
| 圖 4.23 國產蔬菜供貨量關聯度分析-放大圖 | 43 |
| 圖 4.24 國產蔬菜平均價關聯度分析 | 43 |
| 圖 4.25 甘藍-初秋 LA1 供貨量高相關前 5 名品項 | 44 |
| 圖 4.26 甘藍-初秋 LA1 供貨量高相關前 5 名-多品項交易訊息分析..... | 45 |
| 圖 4.27 甘藍-初秋 LA1 供貨量低相關前 5 名品項 | 45 |
| 圖 4.28 甘藍-初秋 LA1 供貨量低相關前 5 名-多品項交易訊息分析..... | 46 |
| 圖 4.29 甘藍-初秋 LA1 平均價高相關前 5 名品項 | 46 |
| 圖 4.30 甘藍-初秋 LA1 平均價高相關前 5 名-多品項交易訊息分析..... | 47 |
| 圖 4.31 甘藍-初秋 LA1 平均價低相關前 5 名品項 | 47 |
| 圖 4.32 甘藍-初秋 LA1 平均價低相關前 5 名-多品項交易訊息分析..... | 48 |

| | |
|---|----|
| 圖 4.33 彰化、雲林、嘉義累積雨量分析..... | 49 |
| 圖 4.34 彰化、雲林、嘉義平均雨量分析..... | 49 |
| 圖 4.35 彰化、雲林、嘉義累積溫度分析..... | 50 |
| 圖 4.36 彰化、雲林、嘉義平均溫度分析..... | 50 |
| 圖 4.37 宜蘭、台中、南投累積雨量分析..... | 50 |
| 圖 4.38 宜蘭、台中、南投平均雨量分析..... | 50 |
| 圖 4.39 宜蘭、台中、南投累積溫度分析..... | 51 |
| 圖 4.40 宜蘭、台中、南投平均溫度分析..... | 51 |
| 圖 5.1 平均價漲跌三分類-前 3 日價格 | 54 |
| 圖 5.2 平均價漲跌三分類-前 10 日價格 | 54 |
| 圖 5.3 累積 30 日溫度關聯度分析..... | 58 |
| 圖 5.4 累積 60 日溫度關聯度分析..... | 60 |
| 圖 5.5 累積 30 日降水量關聯度分析..... | 63 |
| 圖 5.6 累積 60 日降水量關聯度分析..... | 65 |
| 圖 6.1 5-D_LSTM -MSE 折線圖 | 70 |
| 圖 6.2 5-D_LSTM -MSE 折線圖-訓練最後 1000 次 | 70 |
| 圖 6.3 5-D_LSTM 測試集-預測結果 | 70 |
| 圖 6.4 5-D_LSTM 測試集-MSE 折線圖 | 71 |
| 圖 6.5 3-D_LSTM 測試集-預測結果 | 72 |
| 圖 6.6 3-D_LSTM 測試集-MSE 折線圖 | 72 |
| 圖 6.7 10-D_LSTM 測試集-預測結果 | 73 |
| 圖 6.8 10-D_LSTM 測試集-MSE 折線圖 | 73 |
| 圖 6.9 15-D_LSTM 測試集-預測結果 | 73 |
| 圖 6.10 15-D_LSTM 測試集- MSE 折線圖 | 74 |
| 圖 6.11 價格預測混合模型-隨機實驗 1 | 75 |
| 圖 6.12 價格預測混合模型-隨機實驗 2 | 75 |

圖 6.13 價格預測混合模型-隨機實驗 3 76



第一章、緒論

1.1 研究背景

俗話說「民以食為天」。農產品價格的波動影響著我們日常生活。一直以來氣候的變化使得農產品供應量及價格波動，近年來劇烈的氣候變化造成許多異於常態的極端氣候更加劇菜價漲跌。菜價漲跌的影響除了廣泛百姓之外，也包含農產商品供應鏈上的每一個角色，如農民、批發商、餐飲業、中下游農產品攤販。菜價議題一直是學界與公民關心的議題，過去對於農產品價格的分析方法從基本統計模型到機器學習、類神經網路，也可從報章雜誌看見菜價波動報導等，皆顯示農產品價格為台灣重要的社會議題。



數據分析由早期統計模型趨向類神經網路，類神經網路要能夠達到好的效果需要足夠的資料量。近年來政府和民間不斷推動開放資料，從 2011 年 9 月台北市政府資料開放平台上線、2015 年行政院啟動 Open Data 深化應用元年，2016 年成立臺灣資料科學協會，一直到 2016 年台灣獲得開放資料指標（Open Data Index）評比全球第一，由此可見台灣對於資料開放的努力與成效。台灣開放資料能夠擁有這些成效，除了政府支持與推動，許多企業與公民的參與也大大加速發展資料的開放，本研究將開放資料作為主要工具，進行農產品交易訊息視覺化與預測。

本研究針對台北第一果菜市場之國內農產品，挑選全年皆有供貨且年供貨量總和與平均皆為相對大量的甘藍（俗稱高麗菜）作為分析對象。本研究模型加入影響菜價的開放資料作為特徵，預測模型系統使用隨機森林演算法（Random Forest Algorithm）進行特徵分析，最後以類神經網路 LSTM（Long short-term memory）模型進行價格預測實驗。

1.2 研究目的

農產品供應鏈源頭是農民（供貨商）將農產品交給台灣各果菜市場運銷公司，再由全台各批發市場內的拍賣員拍賣，拍賣員會按照政府底價及農產品質量進行拍賣[1]；批發市場周圍或外部也會有行口（中盤商）進行買賣交易，行口的農產品來源可能是批發市場內拍賣而得，但由於交易時間與拍賣衝突因此行口主要會向產地供貨商拉貨，其販售價格除了依據購入價格外也會依循當日拍賣價格。

果菜市場運銷公司和私人行口是交易的源頭，現今許多人對於行口及其他中盤商所販售價格是否公平合理持疑。但此疑慮主要原因是農民與中盤商的資訊雙向不對等，並非兩造之間誰是既得利益者。農民和行口往往會各自站在各自立場：行口對下游販售時壓低價格提高自身競爭力，同時也向農民壓價，有些甚至對農民惡意隱藏市場交易訊息換得更低的買價；農民供貨時會供應超出預期的供貨量或是斷貨，因而提高中盤商出貨與管理成本，有些同樣也會對於

中盤商隱藏產地相關訊息換取更高報酬。無論站在何種角色都各有各的苦衷與難處，更不該二分法的歸咎責任於何方。目前無論是政府及民間企業許多單位正致力於價格的平衡並給予農民保障。解決方法按供應鏈可概分三類：

1. 產地：提高生產技術和生產設備品質，使農產品供貨量不易受環境因素衝擊。
2. 交易過程：提高冷凍倉儲及配送的技術，使農產品較不易受運送過程中敗壞；建立更透明近人的產地種植資訊、交易資訊及氣候環境數據，以減少農民爭先種植同一種品項造成價格崩跌現象、避免因天氣現象造成生產過量或短缺。
3. 消費市場：加強消費者農產品觀念，由於許多成熟的農產品因其生長環境及自然生長造成賣相不佳，但對於飲食健康仍無虞，卻成為淘汰的格外品，因此提升消費者對於格外品的購買意願，也有助於提高供貨量，降低供應鏈內部成本。

廣的來看農產品價格漲跌仍符合市場供需機制，本研究期望利用開放資料建立更平易近人的農產品相關訊息，如農產品歷史供貨量及價格關係視覺化。並且將影響價格漲跌的特徵建立出價格預測模型，使得農民與中盤商於販售農產品前有更便捷的合理資訊來制定價格與交易決策。

本研究應用開放資料於台北第一果菜市場之甘藍價格作為資料特徵數據，其次進行特徵工程包括資料清洗、標記、建構，最後運用隨機森林演算法分析特徵，並以類神經網路 LSTM 建立價格預測模型，本研究目的如下：

1. 資料特徵工程是資料分析的前段工作，且其方法深刻影響分析結果，藉由本研究實驗過程將提出特徵工程方法流程。
2. 建立高效率自動化程式自動抓取最新的開放資料。
3. 提出資料視覺化方法，將大量數值資料轉換成多種圖表，以便於使用者瀏覽資料時可以更快速直觀地觀察資料和統計結果。
4. 搭建類神經網路 LSTM 模型預測甘藍價格。

1.3 研究架構



本研究分為三大部分：第三章說明資料擷取與特徵工程；第四章為資料視覺化；第五、六章呈現特徵分析與價格預測模型，本研究架構模型見圖 1.1，研究架構流程如下：

1. 確定議題範疇：決定研究方向與目標品項。本研究決定目標為預測單一大宗蔬菜：甘藍。
2. 自動化資料擷取與特徵工程：搜集農產品交易開放資料、氣象開放資料及颱風警報資料，並藉由多次試算和調整後提出特徵工程流程。確定特徵工程流程後再將這些方法模組化並設計出自動化擷取程式以便於日後

快速更新資料集。

3. 資料視覺化：將擷取下來的資料繪製成可互動式圖表以便於觀察同時針對農產品供應及交易的特殊現象提出討論。
4. 價格預測模型：運用隨機森林演算法選出對於價格影響力高的特徵資料，再運用 LSTM 模型建置價格預測模型進行訓練與預測實驗以驗證模型成效。

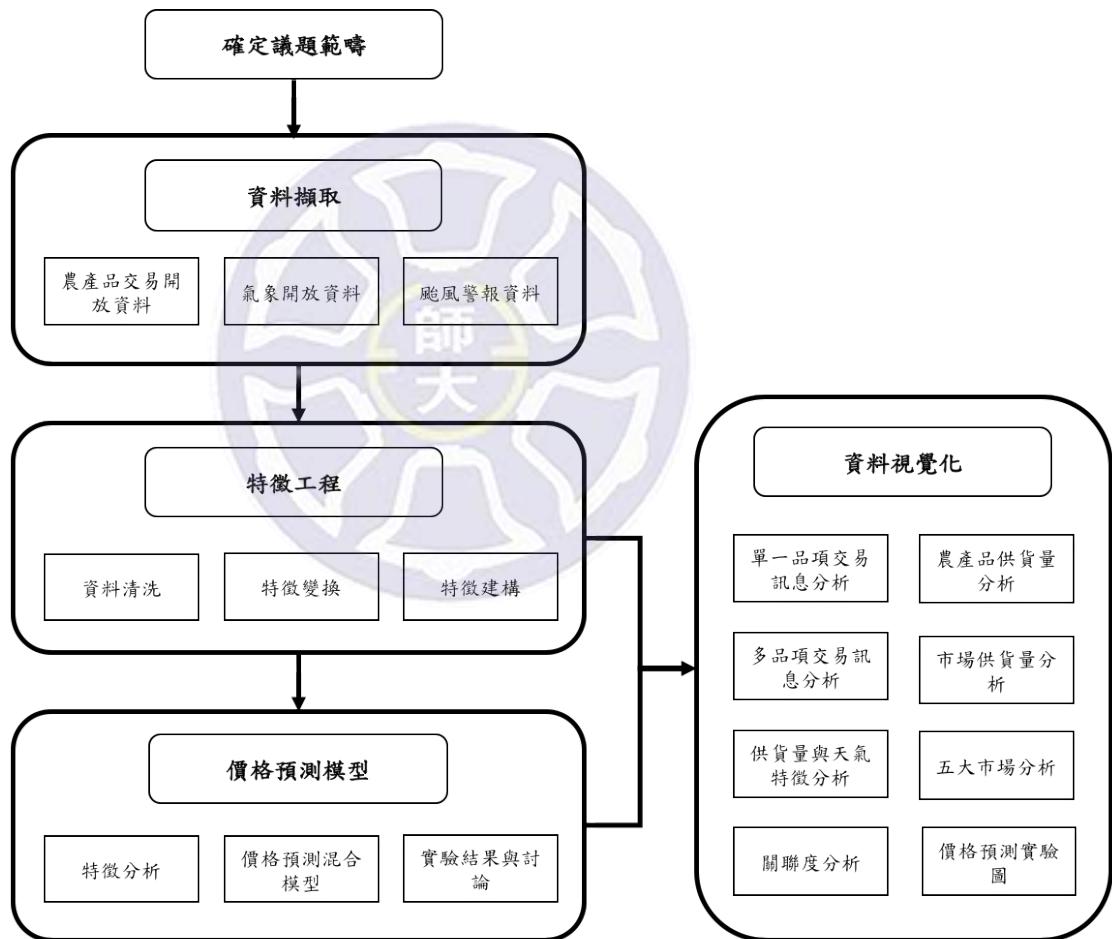


圖 1.1 研究架構圖

第二章、文獻探討

2.1 開放資料

行政院於 2012 年第 3322 次院會決議[2]，政府資料開放，可增進政府施政透明度、提升民眾生活品質，滿足產業界需求，對於各級政府間或各部會間之決策品質均有助益。自 2011 年 9 月臺北市政府資料開放平台上線至今，政府各單位機關及民眾參與使得台灣開放資料於開放資料指標（Open Data Index）於 2016 年評比更獲得全球第一。



2.2 特徵工程

特徵工程是將原始資料轉換成特徵（Features）整個過程的總稱，需要結合該議題專業知識並觀察與分析原始資料集特性進行，進而提高機器學習品質。特徵工程[3]、[4]除了資料除錯、補空值、標準化、正規化之外往往因為資料型態的不同，而需採用不同做法，在此列舉但不限於下列方法：

1. 特徵建構（Feature Construction）：按照該領域的專業知識，將原始某些特徵由人工方式建構出新特徵，如甘藍生長週期範圍約 50 至 60 天 [5]，因此造出新特徵：累積 60 天降水量，藉由建構新的特徵來提高價格預測模型的結果。
2. 特徵轉換（Feature Transform）：若某特徵為連續資料，或精確度到小數點後 n 位，或不同特徵數值的範圍差距甚大，為了提高訓練成效，可

以使用下列方式：四捨五入（ Rounding ）、對數（ Log Transformation ）、分類（ Binning ）：將資料按照其意義分類，如降水量資料該日降水量達降雨規模小雨雨量便指定該值為 1 ；降雨規模大雨雨量便指定該值為 2 ；降水量為 0 指定該值為 0 、 OHE （ One-hot Encoding ）：若該資料包含 n 種值，將其轉換成程度為 n 的向量。

2.3 隨機森林演算法

機器學習中決策樹的分類方法一直是最具代表性的演算法之一。1995 年貝爾實驗室 [6] 提出隨機森林演算法（ Random Forest Algorithm ）。該方法可以視為多個決策樹的組成 [7] ，訓練過程會將訓練資料隨機抽樣再由多個的決策樹將資料進行分類，經過多次抽樣後建造出多顆決策樹。每一顆決策樹為一個弱分類器，再透過母資料群抽樣的資料進行檢驗和評估其分類效力，最終再將每個弱分類器結合成一個強分類器。隨機森林演算法事實上是由眾多弱分類器組合而得的一個強分類器，如此一來可解決一般決策樹分類效果不佳的問題，且可以運用在大量資料上。

2.4 類神經網路之 LSTM

類神經網路（ Artificial Neural Network ），概念源自於訊號在生物神經系統裡傳遞過程而建立的運算模型，依循此概念發展出許多不同類型的類神經網

路，主要結構分為輸入層、隱藏層、輸出層：輸入層用於接收資料的傳入；隱藏層可以為多層，內涵眾多神經元，每個神經元可以視同一個函數，神經元可以彼此相連，運算過程是將輸入及模型內部參數進行運算並輸出，該輸出值可以是下一隱藏層的輸入，也可以是輸出層的輸入，且隱藏層的設計按照不同神經網路模型的功能與特色有所不同；輸出層接收隱藏層的最後一層輸出，並且按照訓練目標輸出結果。

類神經網路至今已發展出各式各樣的模型，本研究之農產品價格預測模型選擇 LSTM 長短期記憶模型，最早由 1997 年提出，其概念由 Recurrent Neural Network (RNN) 遷迴式神經網路[8]演化而來。RNN 優勢為針對具時間序列特性的資料訓練有較佳的結果，如：自然語言分析。相較於 RNN，LSTM 長短期記憶模型加入了「輸入」、「輸出」與「忘記」參數，其中最關鍵為忘記參數，始得 LSTM 同樣是具有記憶功能的類神經網路在訓練時能夠獲得更好的結果，因此本研究挑選 LSTM 長短期記憶模型進行農產品價格預測。

2.5 植物生長理論

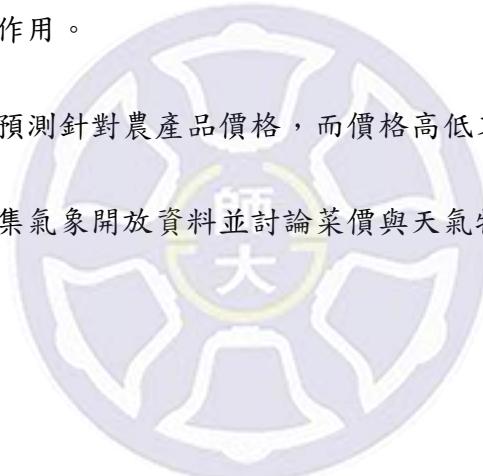
由於植物的生長之基本生理作用包含葉綠體進行光合作用產生並儲存有機物；粒線體和細胞液進行呼吸作用釋放能量；莖內維管束輸送水分與養分，由此可知影響其生長主要因素包含[9][10]：

1. 光照：太陽能是植物生長的主要能量來源，植物透過葉綠體行光合作

用將無機物（二氣化碳、水或硫化氫）生產水和碳水化合物。

2. 溫度：每種作物生長均有其適宜溫度、最低溫度及最高溫度，溫度過低造成代謝異常；過高會抑制光合作用，甚至熱傷害，將使作物停止生長或死亡。作物從生長到收穫期間有一定的累積熱量，稱為「積算溫度」，積算溫度為作物於生長期間超過最低溫度之度數累積。
3. 降水量、濕度：降水量的多寡直接影響植物生長。濕度越低易造成土壤乾硬，造成土壤通氣不良影響植物根部呼吸；空氣中濕度愈大造成降低蒸散作用。

本研究目標為預測針對農產品價格，而價格高低又與生長結果密不可分，因此研究過程將搜集氣象開放資料並討論菜價與天氣特徵之關係。



第三章、資料擷取與特徵工程

3.1 開放資料擷取

本研究首先擷取農產品相關開放資料，使用 Python 之網頁訪問套件 requests 撰寫自動化及多執行緒程式來擷取資料，除了提高資料擷取效率，日後更新資料亦無需人工判斷上一次資料擷取的日期，程式便可自己完成。擷取資料項目及來源見表 3.1。

表 3.1 開放資料來源

| 資料集名稱 | 來源 | 網址 |
|-----------|----------------|---|
| 農產品交易開放資料 | 農產品批發市場 | http://data.coa.gov.tw/Service/OpenData/FromM/FarmTransData.aspx |
| 氣象開放資料 | 中央氣象局：觀測資料查詢系統 | http://e-service.cwb.gov.tw/HistoryDataQuery |
| 颱風警報 | 中央氣象局：歷史颱風 | http://rdc28.cwb.gov.tw |

本章特徵工程綜合整體實驗流程，經過多次測試與不同特徵處理，最終將特徵工程為下列流程：

1. 建立新欄位：由原始資料藉由合併、頗析或按需求額外建立出新欄位。
2. 資料清洗：清理資料集不必要或特定資料，以達到資料整體性、輕便性。
3. 特徵變換：經由多次的實驗與實務探討，將部分原始特徵單獨挑

選出來，並針對其意義與特性進行特徵變換。

4. 建立自動化資料擷取程式：將最終確定的特徵工程方法模組化，

並合併最初資料擷取程式，設計出自動化資料更新程式。

3.2 農產品交易開放資料之特徵工程

下列資料從行政院農業委員會開放資料平台之農產品交易行情 API 擷取下來，資料包含：交易日期、作物代號、作物名稱、市場代號、市場名稱、上價（元/公斤）、中價（元/公斤）、下價（元/公斤）、平均價（元/公斤）、供貨量（公斤）等欄位資料。

表 3.2 農產品交易資料集訊息

| 特徵名稱 | 資料筆數 | 型態 |
|------|---------|---------|
| 交易日期 | 5488638 | object |
| 作物代號 | 5488638 | object |
| 作物名稱 | 5488638 | object |
| 市場代號 | 5488638 | int64 |
| 市場名稱 | 5488638 | object |
| 交易量 | 5488638 | float64 |
| 上價 | 5488638 | float64 |
| 中價 | 5488638 | float64 |
| 下價 | 5488638 | float64 |
| 平均價 | 5488638 | float64 |

農產品交易開放資料擷取結果見表 3.2，欄位說明如下：

1. 供貨量：當日農產品於該市場供貨量，單位為公斤。
2. 上價：當日該農產品於該市場中最高價格 20%，加權平均計算得之，單位為元/公斤。

3. 中價：當日該農產品於該市場扣掉價格最高及最低各 20%後剩餘的 60%，再加權平均計算得之。
4. 下價：當日該農產品於該市場中最低價格 20%，加權平均計算得之，單位為元/公斤。
5. 市場名稱：市場代碼表[11]見圖 3.1。

行政院農業委員會開放資料平台 - API 介接說明書

7. 市場代碼表

| 蔬菜 | | 水果 | | 花卉 | |
|-----|-----|-----|-----|-----|----|
| 109 | 台北一 | 109 | 台北一 | 105 | 台北 |
| 104 | 台北二 | 104 | 台北二 | 400 | 台中 |
| 220 | 板橋區 | 220 | 板橋區 | 514 | 彰化 |
| 241 | 三重區 | 241 | 三重區 | 700 | 台南 |
| 338 | 桃農 | 338 | 桃農 | 800 | 高雄 |
| 400 | 台中 | 400 | 台中 | | |
| 420 | 豐原 | 420 | 豐原 | | |
| 540 | 南投 | 423 | 東勢 | | |
| 512 | 永靖 | 600 | 嘉義 | | |
| 514 | 溪湖 | 800 | 高雄 | | |
| 648 | 西螺 | 830 | 鳳山區 | | |
| 800 | 高雄 | 930 | 台東 | | |
| 830 | 鳳山區 | 260 | 宜蘭 | | |
| 900 | 屏東 | 540 | 南投 | | |
| 930 | 台東 | | | | |
| 260 | 宜蘭 | | | | |
| 950 | 花蓮 | | | | |

圖 3.1 農產品市場代碼表

3.2.1 建立新欄位

我們將農產品開放資料擷取下來並觀察內部數據見圖 3.2，再進行特徵工程。首先建立新欄位：作物名稱代號、Year、Month、Day、交易日期_西元、GetTime。

| | 上價 | 下價 | 中價 | 交易日期 | 交易量 | 作物代號 | 作物名稱 | 市場代號 | 市場名稱 | 平均價 |
|---|-------|------|-------|-----------|--------|------|---------|------|------|-------|
| 0 | 16.2 | 9.3 | 11.7 | 106.01.01 | 1417.0 | 11 | 椰子 | 104 | 台北二 | 12.1 |
| 1 | 130.0 | 55.3 | 80.0 | 106.01.01 | 5667.0 | 22 | 棗子 | 104 | 台北二 | 85.0 |
| 2 | 142.4 | 63.4 | 100.9 | 106.01.01 | 739.0 | 31 | 釋迦 | 104 | 台北二 | 101.7 |
| 3 | 102.2 | 47.2 | 63.0 | 106.01.01 | 935.0 | 32 | 釋迦-鳳梨釋迦 | 104 | 台北二 | 67.7 |
| 4 | 170.3 | 83.8 | 122.0 | 106.01.01 | 798.0 | 45 | 草莓 | 104 | 台北二 | 124.0 |

圖 3.2 農產品開放資料集

1. 作物名稱代號：建立農產品交易資料的鍵值。由於原始的「作物代號」並非唯一值，即同一農產品品項有更細的品種，如「作物名稱」蘿蔔-研仔（白蘿蔔）有兩個作物代號：SA31、SA32 兩項品種；或是隨著時間及市場變化，使得部分農產品的代號更新。

由於此欄位是農產品交易開放資料集的主要欄位，攸關於日後篩選品項的精準度及分析的可信度，因此必須建立出農產品資料集的鍵值。方法為合併「作物名稱」、「作物代號」兩欄位，例如本研究目標品項為甘藍，其作物名稱為「甘藍-初秋」；作物代號為「LA1」，合併結果為「甘藍-初秋 LA1」。

2. Year：該筆交易日期之年份。由於原始「交易日期」的年份計算方式為民國，為了統一格式以便於未來與其他特徵資料集合併及篩選日期，因此頗析出交易日期的年份，並轉換成西元格式。

3. Month：交易日期的月份，頗析出「交易日期」之月份。
4. Day：交易日期的日期。頗析出「交易日期」之日期。
5. 交易日期_西元：該筆資料交易日期。合併「Year」、「Month」、「Day」

三欄位，並統一格式由符號「 - 」分隔年、月、日，且月份和日期固定

格式為兩位數：YY-MM-DD（年-月-日），若不到兩位數補0，例：

2017-04-07。

6. GetTime：資料擷取時間紀錄。由於資料會隨著時間更新，且農產品開

放資料更新時間為每日4次[12]，又不同市場更新時間不一定相同，為

了避免日後於資料更新後造成同日期資料重複、混淆，因此紀錄資料

擷取的時間，以便於日後遇到資料重複時，可以判別資料新舊。欄位

建立結果見圖3.3。

| GetTime | 上價 | 下價 | 中價 | 交易日期 | 交易量 | 作物代號 | 作物名稱 | 市場代號 | 市場名稱 | 平均價 | 作物名稱代號 | Year | Month | YearMonth | Day | 交易日期_西元 |
|---------------------|-------|------|------|-----------|--------|------|---------|------|------|------|------------|------|-------|-----------|-----|------------|
| 2018-02-19 02:58:58 | 15.0 | 7.5 | 10.1 | 101.01.01 | 1700.0 | 11 | 椰子 | 104 | 台北二 | 10.5 | 椰子 11 | 2012 | 01 | 2012-01 | 01 | 2012-01-01 |
| 2018-02-19 02:58:58 | 89.4 | 16.0 | 44.6 | 101.01.01 | 9888.0 | 22 | 棗子 | 104 | 台北二 | 47.8 | 棗子 22 | 2012 | 01 | 2012-01 | 01 | 2012-01-01 |
| 2018-02-19 02:58:58 | 66.6 | 22.9 | 46.1 | 101.01.01 | 1011.0 | 31 | 釋迦 | 104 | 台北二 | 45.6 | 釋迦 31 | 2012 | 01 | 2012-01 | 01 | 2012-01-01 |
| 2018-02-19 02:58:58 | 85.2 | 24.4 | 34.4 | 101.01.01 | 1558.0 | 32 | 釋迦-鳳梨釋迦 | 104 | 台北二 | 42.6 | 釋迦-鳳梨釋迦 32 | 2012 | 01 | 2012-01 | 01 | 2012-01-01 |
| 2018-02-19 02:58:58 | 110.0 | 41.8 | 63.3 | 101.01.01 | 1522.0 | 45 | 草莓 | 104 | 台北二 | 68.3 | 草莓 45 | 2012 | 01 | 2012-01 | 01 | 2012-01-01 |

圖3.3 農產品資料-建立新欄位結果

3.2.2 刪除市場休息資料

由於果菜市場逢星期一或特定節日會休息（圖3.4），休息日的價格及供貨量數值資料皆為0值、「作物名稱」欄位值為「休市」，但是2015年以前沒有「休市」值，資料數據會跳過休市日，因此無法統一由此欄位篩選出休市日期資料，遂統一將休息日都刪除，使資料集裡不存在的日期就是休市日期。

| GetTime | 上價 | 下價 | 中價 | 交易日期 | 交易量 | 作物代號 | 作物名稱 | 市場代號 | 市場名稱 | 平均價 | 作物名稱代號 | Year | Month | YearMonth | Day | 交易日期_西元 |
|---------------------|-----|-----|-----|-----------|-----|------|------|------|------|-----|---------|------|-------|-----------|-----|------------|
| 2018-02-19 03:34:08 | 0.0 | 0.0 | 0.0 | 104.01.01 | 0.0 | rest | 休市 | 423 | 東勢鎮 | 0.0 | 休市 rest | 2015 | 01 | 2015-01 | 01 | 2015-01-01 |
| 2018-02-19 03:34:16 | 0.0 | 0.0 | 0.0 | 104.01.04 | 0.0 | rest | 休市 | 423 | 東勢鎮 | 0.0 | 休市 rest | 2015 | 01 | 2015-01 | 04 | 2015-01-04 |
| 2018-02-19 03:34:16 | 0.0 | 0.0 | 0.0 | 104.01.04 | 0.0 | rest | 休市 | 512 | 永靖鄉 | 0.0 | 休市 rest | 2015 | 01 | 2015-01 | 04 | 2015-01-04 |
| 2018-02-19 03:34:16 | 0.0 | 0.0 | 0.0 | 104.01.04 | 0.0 | rest | 休市 | 648 | 西螺鎮 | 0.0 | 休市 rest | 2015 | 01 | 2015-01 | 04 | 2015-01-04 |
| 2018-02-19 03:34:16 | 0.0 | 0.0 | 0.0 | 104.01.04 | 0.0 | rest | 休市 | 950 | 花蓮市 | 0.0 | 休市 rest | 2015 | 01 | 2015-01 | 04 | 2015-01-04 |

圖 3.4 農產品資料-市場休息資料

3.2.3 轉換欄位名稱為英文

為了後續資料操作方便，將欄位名稱轉換成英文，轉換方式見表 3.3。

表 3.3 農產品資料-欄位中英文對照表

| 原始中文名稱 | 英文名稱 |
|---------|-----------------|
| 上價 | BestPrice |
| 下價 | GoodPrice |
| 中價 | BetterPrice |
| 交易日期_西元 | TransactionDate |
| 交易量 | Amount |
| 作物代號 | Item_ID |
| 作物名稱 | Item_Name |
| 市場代號 | Market_ID |
| 市場名稱 | Market_Name |
| 平均價 | AveragePrice |
| 作物名稱代號 | Item_NameID |

3.2.4 刪除相依欄位

由於資料量太龐大：六年的資料量為五百多萬筆，每多一欄會增加大量儲存空間，為了節省儲存空間，因此刪除不必要欄位。如：Year、Month、Day 與 TransactionDate 為相依資料；Item_Name、Item_ID 與 Item_NameID 為相依資料欄；市場名稱與 Market_ID 其意義相同，因此保留 Market_Name 刪掉 Market_ID。未來若有需要這些欄位，如針對年份或月份獨立篩選資料集，只需要頗析 TransactionDate 欄位便可獲得。刪除不必要欄位後結果見圖 3.5。

| TransactionDate | Item_NameID | Market_Name | Amount | BestPrice | BetterPrice | GoodPrice | AveragePrice | GetTime |
|-----------------|-------------|-------------|--------|-----------|-------------|-----------|--------------|---------------------|
| 2012-01-01 | 椰子 11 | 台北二 | 1700.0 | 15.0 | 10.1 | 7.5 | 10.5 | 2018-02-19 02:58:58 |
| 2012-01-01 | 棗子 22 | 台北二 | 9888.0 | 89.4 | 44.6 | 16.0 | 47.8 | 2018-02-19 02:58:58 |
| 2012-01-01 | 釋迦 31 | 台北二 | 1011.0 | 66.6 | 46.1 | 22.9 | 45.6 | 2018-02-19 02:58:58 |
| 2012-01-01 | 釋迦-鳳梨釋迦 32 | 台北二 | 1558.0 | 85.2 | 34.4 | 24.4 | 42.6 | 2018-02-19 02:58:58 |
| 2012-01-01 | 草莓 45 | 台北二 | 1522.0 | 110.0 | 63.3 | 41.8 | 68.3 | 2018-02-19 02:58:58 |

圖 3.5 農產品資料-最終格式

原始資料集一共有 11 個欄位（圖 3.6），5488638 筆資料，約佔儲存空間 502.5 MB。

```
Int64Index: 5488638 entries, 0 to 984221
Data columns (total 11 columns):
GetTime      object
上價         float64
下價         float64
中價         float64
交易日期     object
交易量        float64
作物代號    object
作物名稱    object
市場代號    int64
市場名稱    object
平均價        float64
dtypes: float64(5), int64(1), object(5)
memory usage: 502.5+ MB
```

圖 3.6 農產品資料-整理前訊息

經過上述特徵工程與資料清洗後（圖 3.7），資料欄位剩下 9 欄，共 5484666 筆數據，約佔儲存空間 376.6 MB，前後相比較，減少 3972 筆並節省空間約 125 MB。

```
RangeIndex: 5484666 entries, 0 to 5484665
Data columns (total 9 columns):
TransactionDate      object
Item_NameID          object
Market_Name           object
Amount                float64
BestPrice             float64
BetterPrice            float64
GoodPrice              float64
AveragePrice          float64
GetTime               object
dtypes: float64(5), object(4)
memory usage: 376.6+ MB
```

圖 3.7 農產品資料-整理後訊息

3.2.5 建構日期特徵與市場休息特徵

由於農產品交易資料只有在 2015 年以後才有再開放資料裡標記市場休息日，在此之前資料數值會跳過休息日，先前於「3.2.2 刪除市場休息資料」統一刪除市場休息資料，因此需獨立將市場休息特徵建構出來，步驟如下：

1. 建立「日期特徵」資料集，資料範圍從 2012 年至 2018 年（由於資料會不斷更新因此建立至 2018 年底），欄位包含 Year (年)、Month (月)、WeekDay (星期)、Date (日期)，結果見圖 3.8。

| | Year | Month | WeekDay | Date |
|---|------|-------|---------|------------|
| 0 | 2012 | 1 | 7 | 2012-01-01 |
| 1 | 2012 | 1 | 1 | 2012-01-02 |
| 2 | 2012 | 1 | 2 | 2012-01-03 |
| 3 | 2012 | 1 | 3 | 2012-01-04 |
| 4 | 2012 | 1 | 4 | 2012-01-05 |

圖 3.8 日期資料集

2. 將農產品資料集按日期合併進「日期特徵」，並且判斷該日供貨量與價格是否有空值，再建立「RestMark」欄位，若有空值則賦值「空欄」，結果見圖 3.9，如此一來便可挑選並建立出市場休息特徵。

| TransactionDate | Item_NameID | Market_Name | Amount | BestPrice | BetterPrice | GoodPrice | AveragePrice | Rest_Mark | Year | Date |
|-----------------|-------------|-------------|----------|-----------|-------------|-----------|--------------|-----------|------|-------|
| 2012-01-01 | 甘藍-初秋 LA1 | 台北一 | 124858.0 | 39.8 | 35.8 | 30.9 | 35.6 | None | 2012 | 01-01 |
| 2012-01-02 | 甘藍-初秋 LA1 | 台北一 | Nan | Nan | Nan | Nan | Nan | 空欄 | 2012 | 01-02 |
| 2012-01-03 | 甘藍-初秋 LA1 | 台北一 | 173808.0 | 36.7 | 32.7 | 21.5 | 31.3 | None | 2012 | 01-03 |
| 2012-01-04 | 甘藍-初秋 LA1 | 台北一 | 144654.0 | 35.7 | 31.0 | 21.8 | 30.1 | None | 2012 | 01-04 |
| 2012-01-05 | 甘藍-初秋 LA1 | 台北一 | 145523.0 | 33.0 | 28.8 | 21.9 | 28.2 | None | 2012 | 01-05 |
| 2012-01-06 | 甘藍-初秋 LA1 | 台北一 | 150214.0 | 27.3 | 23.1 | 17.5 | 22.8 | None | 2012 | 01-06 |
| 2012-01-07 | 甘藍-初秋 LA1 | 台北一 | 183073.0 | 25.3 | 19.7 | 13.4 | 19.6 | None | 2012 | 01-07 |
| 2012-01-08 | 甘藍-初秋 LA1 | 台北一 | 145742.0 | 30.7 | 26.4 | 19.8 | 25.9 | None | 2012 | 01-08 |
| 2012-01-09 | 甘藍-初秋 LA1 | 台北一 | Nan | Nan | Nan | Nan | Nan | 空欄 | 2012 | 01-09 |

圖 3.9 合併農產品交易與日期資料集結果

3. 休市特徵變換：由於市場休息日對於現實交易的影響可能性很多，根據商業周刊「台北果菜市場 20 天休市 11 天菜價崩跌」[11]：
- 「台北農產運銷公司農曆春節自 2 月 16 日休市 5 天，2 月 24 日至 26 日又休 3 天，致使年後蔬菜到貨爆量、拍賣價暴跌。」
- 因此將休市特徵「Rest_Mark」進行特徵變換，變換方式見表 3.4，結果見圖 3.10。

表 3.4 市場休息特徵變換結果

| 特徵名稱 | 欄位說明 | 欄位資料意義 |
|------------------|-----------|---------------|
| Rest_Mark | 當天市場是否休息 | |
| YesterDayRest | 前一天市場是否休息 | |
| TomorrowRest | 隔天市場是否休息 | |
| YesterDayRestCNT | 前一天連休幾日 | |
| TomorrowRestCNT | 隔天連休幾日 | 若連休 x 日，該值為 x |

| | TransactionDate | Rest_Mark | YesterDayRest | TomorrowRest | YesterDayRestCNT | TomorrowRestCNT |
|------|-----------------|-----------|---------------|--------------|------------------|-----------------|
| 1498 | 2016-02-07 | 0.0 | 0.0 | 1.0 | 0.0 | 5.0 |
| 1499 | 2016-02-08 | 1.0 | 0.0 | 1.0 | 0.0 | 4.0 |
| 1500 | 2016-02-09 | 1.0 | 1.0 | 1.0 | 1.0 | 3.0 |
| 1501 | 2016-02-10 | 1.0 | 1.0 | 1.0 | 2.0 | 2.0 |
| 1502 | 2016-02-11 | 1.0 | 1.0 | 1.0 | 3.0 | 1.0 |
| 1503 | 2016-02-12 | 1.0 | 1.0 | 0.0 | 4.0 | 0.0 |
| 1504 | 2016-02-13 | 0.0 | 1.0 | 0.0 | 5.0 | 0.0 |
| 1505 | 2016-02-14 | 0.0 | 0.0 | 1.0 | 0.0 | 2.0 |
| 1506 | 2016-02-15 | 1.0 | 0.0 | 1.0 | 0.0 | 1.0 |
| 1507 | 2016-02-16 | 1.0 | 1.0 | 0.0 | 1.0 | 0.0 |
| 1508 | 2016-02-17 | 0.0 | 1.0 | 0.0 | 2.0 | 0.0 |

圖 3.10 休市特徵資料集

3.3 氣象開放資料之特徵工程

中央氣象局觀測資料查詢系統提供全台灣各地氣象測站觀測資料[13]，分為日報表、月報表、年報表。日報表紀錄每小時為單位之測站數據；月報表紀錄每天測站數據；年報表則按該年每月方式紀錄。其數據內容按照特性分成平均值和累積值紀錄，如風速、風向以平均值計算；降水量、降水時數、日照時數、全天空日射量以累積值計算。由於日報表紀錄過於詳細，年報表又過於籠

統，因此本研究擷取紀錄每日測站數據的月報表，以配合農產品交易開放資料每日更新的特性。

擷取氣象開放資料前，須先搜集整理各地測站編號，測站分成三種類型，按照測站編號前兩碼分為 46、C0、C1。46 測站為中央氣象局地面氣象站測站，包含全部測量數據。C 開頭測站為中央氣象局自動氣象站（C0）和雨量站（C1），C0 測站包含測站氣壓、氣溫、相對溼度、風速、風向、降水量；C1 測站僅紀錄降水量。由此可知並非每一個氣象觀測站的資料量皆相同，且測站遇故障會有缺值。因此按照測站所在地及屬性建立測站資料集，過程中建立新欄位「InnerIsland」用來紀錄該測站是否位於台灣本島，如在本島內賦值為 1，反之為 0，氣象測站資料及欄位說明見表 3.5，氣象測站資料集見圖 3.11。

表 3.5 氣象測站代碼欄位說明

| 欄位名稱 | 欄位說明 |
|-------------|--------|
| StationCode | 測站編號 |
| ChName | 測站所處地點 |
| Area | 測站所處城市 |
| InnerIsland | 是否位於島內 |

| | StationCode | ChName | Area | InnerIsland |
|---|-------------|--------|------|-------------|
| 0 | 466880 | 板橋 | 新北市 | 1 |
| 1 | 466900 | 淡水 | 新北市 | 1 |
| 2 | 466910 | 鞍部 | 臺北市 | 1 |
| 3 | 466920 | 臺北 | 臺北市 | 1 |
| 4 | 466930 | 竹子湖 | 臺北市 | 1 |

圖 3.11 氣象測站代碼資料集

獲得測站編號後，便可擷取氣象資料，擷取結果之六年資料量約一百萬筆，擷取結果見表 3.6。

表 3.6 氣象開放資料集訊息

| 特徵名稱 | 型態 | 特徵名稱 | 型態 |
|-------------|---------|---------------|--------|
| MeasureTime | object | 最小相對溼度 | object |
| StationCode | object | 最小相對溼度時間 | object |
| Year | object | 最低氣溫 | object |
| Month | object | 最低氣溫時間 | object |
| Day | object | 最高氣溫 | object |
| YearMonth | object | 最高氣溫時間 | object |
| ChName | object | 測站最低氣壓 | object |
| EnName | object | 測站最低氣壓時間 | object |
| Area | object | 測站最高氣壓 | object |
| InnerIsland | int64 | 測站最高氣壓時間 | object |
| 觀測時間 | object | 最大陣風 | object |
| 降水量 | object | 最大陣風風向 | object |
| 氣溫 | object | 最大陣風風速時間 | object |
| 相對溼度 | object | 降水時數 | object |
| 風速 | object | 10 分鐘最大降水量 | object |
| 測站氣壓 | object | 10 分鐘最大降水起始時間 | int64 |
| 風向 | object | 一小時最大降水量 | object |
| A 型蒸發量 | object | 一小時最大降水量起始時間 | object |
| 能見度 | object | 全天空日射量 | object |
| 海平面氣壓 | object | 日照時數 | object |
| 露點溫度 | float64 | 日照率 | object |

3.3.1 特徵工程

氣象資料之特徵工程步驟和農產品開放資料型態相似，包括觀察原始數據

(圖 3.12)、建立新欄位、轉換欄位名稱為英文、資料清洗等，步驟如下。

| 觀測時間 | 測站氣壓 | 海平面氣壓 | 測站最高氣壓時間 | 測站最低氣壓時間 | 測站最高氣溫時間 | 測站最低氣溫時間 | 最高氣溫 | 最高氣溫時間 | … | 10分鐘最大降水量起始時間 | 一小時最大降水量 | 日大降水量 | 日時數 | 日照時數 | 全天空日射量 | 能見度 | A型蒸發量 | MeasureTime | StationCode |
|------|--------|--------|----------------------------|-----------------|-----------------|---------------|----------------|------------------|-----|------------------|----------|-------|-------|------|--------|-----|---------|-------------|-------------|
| 0 1 | 1006.8 | 1008 | 1008.8 2011-06-01 22:47 | 1005.6 04:46 | 2011-06-01 26.0 | 32.1 11:12 | 2011-06-01 ... | 2011-06-01 15:07 | 3.0 | 2011-06-01 14:46 | 1.2 | 8.9 | 8.31 | 7.6 | -1.4 | | 2011-06 | 466880 | |
| 1 2 | 1008.5 | 1009.7 | 1010.2 2011-06-02 09:23 | 1006.6 02:25 | 2011-06-02 25.3 | 29.3 12:46 | 2011-06-02 ... | 2011-06-02 09:35 | 7.5 | 2011-06-02 08:55 | 1.1 | 8.1 | 9.50 | 11.5 | -2.3 | | 2011-06 | 466880 | |
| 2 3 | 1008.8 | 1010 | 1010.2 2011-06-03 09:02 | 1007.5 16:51 | 2011-06-03 25.5 | 29.3 14:11 | 2011-06-03 ... | 2011-06-03 07:40 | T | 2011-06-03 07:40 | 0.2 | 1.5 | 6.27 | 7.4 | -0.7 | | 2011-06 | 466880 | |
| 3 4 | 1007.4 | 1008.6 | 1009.0 2011-06-04 07:01 | 1005.7 17:14 | 2011-06-04 27.7 | 32.0 11:57 | 2011-06-04 ... | 2011-06-04 06:10 | T | 2011-06-04 06:10 | 1.7 | 12.5 | 14.85 | 5.3 | 2.5 | | 2011-06 | 466880 | |
| 4 5 | 1007.0 | 1008.2 | 1008.2 2011-06-05 21:59 | 1005.9 05:02 | 2011-06-05 29.0 | 34.5 12:00 | 2011-06-05 ... | Nan | 0.0 | Nan | 8.4 | 61.9 | 23.58 | 10.3 | 4.6 | | 2011-06 | 466880 | |

5 rows × 34 columns

圖 3.12 氣象資料集

1. 建立新欄位：「 MeasureTime 」：該筆測站資料日期，統一格式：YY-MM-DD 。
2. 轉換欄位名稱為英文，轉換方式見表 3.7 。

表 3.7 氣象資料-欄位中英文對照表

| 原始中文名稱 | 英文名稱 |
|--------|-------------|
| 測站氣壓 | Pressure |
| 氣溫 | Temperature |
| 相對溼度 | Humidity |
| 降水量 | Rain |

3. 資料清洗：按照「 InnerIsland 」資料，刪除位於島外的測站。按上述特徵工程結果見圖 3.13 。

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1055306 entries, 0 to 1055305
Data columns (total 37 columns):
觀測時間           1055306 non-null int64
Pressure          718198 non-null object
海平面氣壓         74678 non-null object
測站最高氣壓       708351 non-null object
測站最高氣壓時間   684685 non-null object
測站最低氣壓       708351 non-null object
測站最低氣壓時間   683132 non-null object
Temperature        731743 non-null object
最高氣溫           708351 non-null object
最高氣溫時間       684062 non-null object
最低氣溫           708351 non-null object
最低氣溫時間       684296 non-null object
露點溫度          74678 non-null object
Humidity          608482 non-null object
最小相對溼度       707982 non-null object
最小相對溼度時間   574456 non-null object
風速              731743 non-null object
風向              731743 non-null object
最大陣風           708351 non-null object
最大陣風風向       708351 non-null object
最大陣風風速時間   684102 non-null object
Rain              1054534 non-null object
降水時數          72563 non-null object
10分鐘最大降水量    74678 non-null object
10分鐘最大降水起始時間 32211 non-null object
一小時最大降水量    74678 non-null object
一小時最大降水量起始時間 32218 non-null object
日照時數          117743 non-null object
日照率            74678 non-null object
全天空日射量       74678 non-null object
能見度            68462 non-null float64
A型蒸發量         65895 non-null object
MeasureTime        1055306 non-null object
StationCode        1055306 non-null object
ChName            1055306 non-null object
Area              1055306 non-null object
InnerIsland        1055306 non-null int64
dtypes: float64(1), int64(2), object(34)
memory usage: 297.9+ MB

```

圖 3.13 氣象資料-整理後訊息

3.3.2 獨立各縣市天氣特徵

由於氣象資料集包含許多本研究當前使用不到的氣象觀測數據，且資料儲存型式為每一測站數據，全台灣有 500 多個測站，使得每一筆資料僅能表示一個城市裡的某一個區域，實務上從 500 多個測站裡挑選目標測站進行實驗也不太實際，因此需要進一步分析，流程如下：

1. 挑選影響植物生長最大的特徵：溫度、降水量。

2. 資料異常值處理：測站數據值若為 T 表示雨跡（小於 0.1mm）、X 則為故障[12]，本研究於此針對兩項異常值處理：T 由數值 0.05 取代，表示有降下少量的雨；X 直接空值取代，視同捨棄。由此將所有數據轉為非文字型態，以便於下一步計算平均。
3. 整合同一城市所有測站數據：合併相同城市特徵數據並計算平均，建立出台灣 19 縣市（僅含台灣本島內的氣象測站）的天氣特徵資料（圖 3.14）。

| 彰化縣 Temperature | 雲林縣 Temperature | 嘉義市 Temperature | 南投縣 Temperature | 臺南市 Temperature | 高雄市 Temperature | 屏東縣 Temperature | 花蓮縣 Temperature | 宜蘭縣 Temperature | 臺東縣 Temperature | MeasureTime |
|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|-------------|
| 16.470588 | 15.800 ... | 16.5 | 9.03 | 17.033333 | 17.585714 | 18.325000 | 13.585 | 13.47 | 14.9250 | 2012-01-01 |
| 16.017647 | 15.150 ... | 15.6 | 7.85 | 15.833333 | 16.157143 | 17.466667 | 12.735 | 12.46 | 13.9000 | 2012-01-02 |
| 16.605882 | 15.675 ... | 16.6 | 9.29 | 16.950000 | 17.185714 | 18.616667 | 15.290 | 14.32 | 15.8875 | 2012-01-03 |
| 12.423529 | 12.675 ... | 13.2 | 7.87 | 14.000000 | 15.142857 | 16.616667 | 11.575 | 10.13 | 13.5875 | 2012-01-04 |
| 12.305882 | 12.200 ... | 12.3 | 7.52 | 12.666667 | 13.800000 | 15.716667 | 11.470 | 10.55 | 12.9250 | 2012-01-05 |

圖 3.14 各縣市天氣特徵-溫度

4. 觀察空值並找尋適當填補方法：整合同一城市測站天氣特徵時發現部分城市包含缺值，如新竹市：肇因起初該市僅有一個測站，因此若該測站壞掉即造成新竹市當日無氣象數據。但是丟棄十分可惜，因此將各城市天氣數值繪成折線圖，再逐步觀察挑出天氣變化最相似的城市取代缺失值。例如：彰化縣溫度有缺值，首先繪出全部縣市溫度折線圖，並逐一比對鄰近城市溫度，再挑選溫度波形相似度最高的城市。繪圖範例（以彰化縣為例）：首先繪製全縣市溫度折線圖（圖 3.15）。

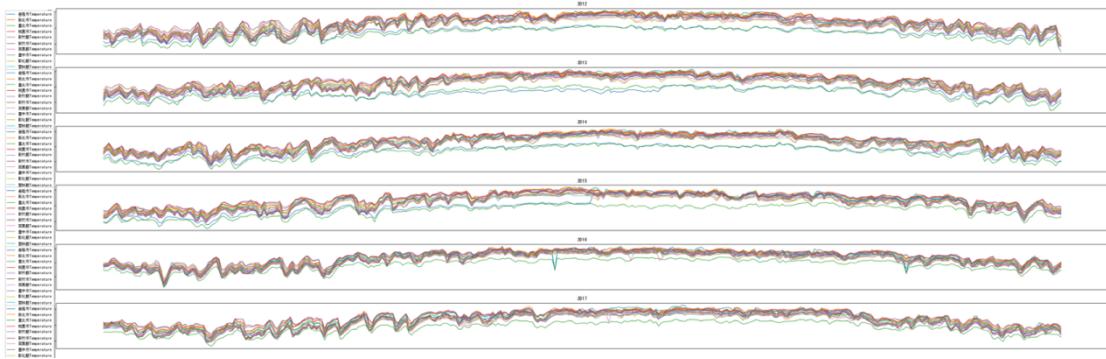


圖 3.15 溫度折線圖-全部縣市

逐步篩選出溫度波形最相似城市：臺中市（圖 3.16）。

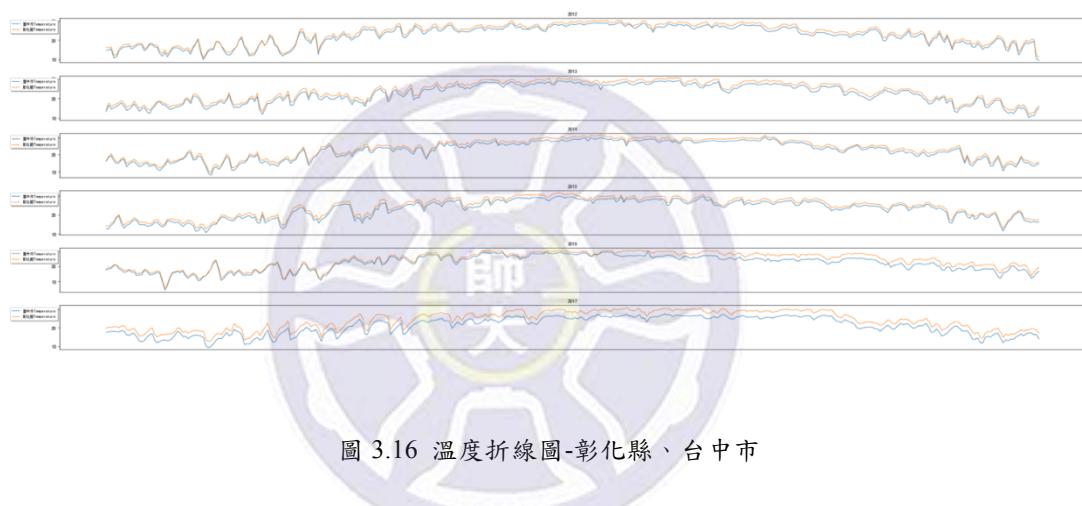


圖 3.16 溫度折線圖-彰化縣、台中市

經過逐一比對，挑選出溫度、降水量兩特徵。並決定出下列城市缺值

修復規則：（含缺值的城市：補缺值的城市）

1. 新竹市：新竹縣。
2. 苗栗縣：新竹縣和臺中市之平均。
3. 彰化縣：臺中市。
4. 雲林縣：嘉義縣。

3.4 建立自動化資料擷取程式

針對農產品交易開放資料及氣象開放資料這兩份資料集，為了方便日後更新資料，因此重新建構原先資料擷取程式，建立自動化程式，使得日後更新時無需仰賴人工檢視上一次更新時間，並且加入多執行緒訪問網頁功能，提高資料擷取的效率。

1. 特徵工程模組化：綜合上述特徵工程，將上述步驟之程式模組化，以利於新進資料清洗與整理。
2. 設計自動化資料擷取程式：判斷前一次爬蟲日期，並接續上次結束日期開始擷取資料，如此一來新進資料的過程無需人工介入，便可以直接更新。
3. 資料檢驗：為了確定自動更新後資料的正確，於是建立檢驗欄位以判斷更新後的資料集是否有重複資料。建立方法為合併日期與該資料集鍵值，如新欄位有重複值，表示同一天有重複資料，發生此情形會按照「GetTime」判斷該筆資料新舊，並刪掉舊的資料，僅保留新資料，檢驗欄位按資料集分為二種：
 - i. 農產品批發價：合併交易日期、品項 ID 和交易市場三欄位。
 - ii. 中央氣象局觀測資料：合併觀測日期及測站編號兩欄位。

第四章、資料視覺化

由於本研究兩資料集擁有龐大資料量，藉由肉眼觀察數據只能看到數筆片段數據。但是如此龐大的資料量且具有時序性，若能細心從中觀察，相信能夠獲得高價值的訊息，應用層面除了提供農民及盤商特定農產品行情、交易現情局勢來執行決策，抑或是找出有趣的現象進行深度研究，因此本研究針對農產品批發價，使用資料視覺化將龐大且繁雜的數據轉化為直觀圖表，並且具有互動式特性，同時設計便利的視覺化模組，以便於觀察各式不同農產品品項。

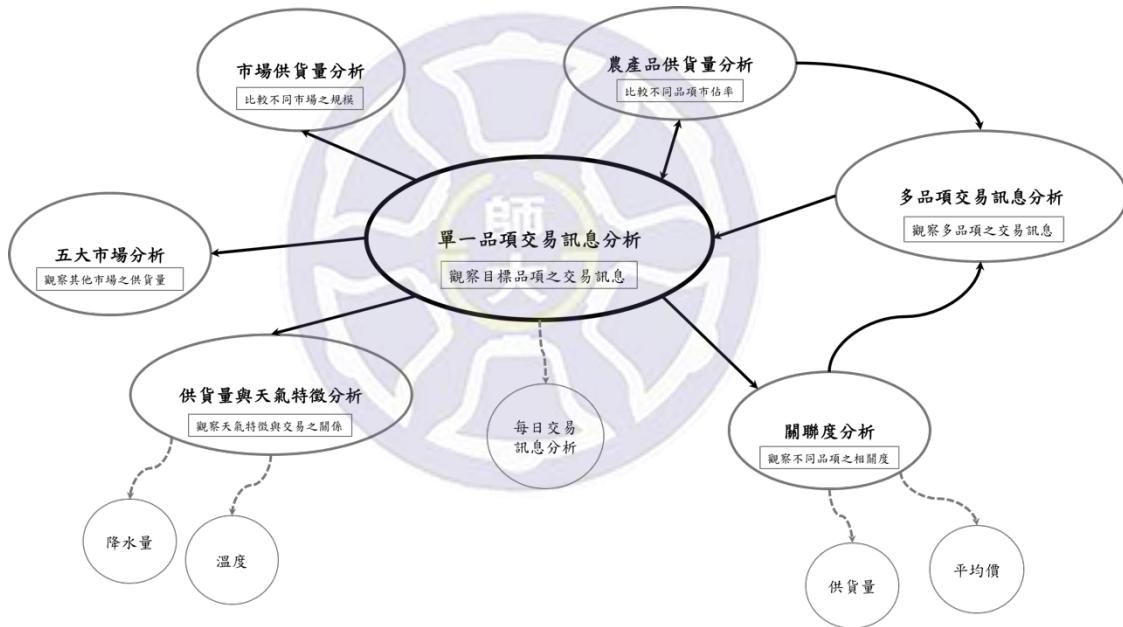


圖 4.1 資料視覺化流程圖

本研究繪圖方法為 Python 之 pycharts 套件，其特色為圖形由網頁呈現且具互動性，可以自由縮放、選取或取消選取標籤以利於觀察圖表，生成的圖片為網頁格式，攜帶方便也具有結合進網頁或 APP 的特性。部分資料視覺化前會先針對資料進行篩選及先前計算，使得肉眼能夠更直觀地觀察數值背後的訊息，

本章數據皆統計自 2012 年起 2017 迄，一共分為七大類型（圖 4.1），本章節僅擷取部分圖表，其餘可見附錄。

1. 農產品供貨量分析：加總各農產品分年度供貨量，分析個別農產品所佔總交易量比例。
2. 市場供貨量分析：加總各市場分年度供貨量，分析全台灣各市場交易量。
3. 單一品項交易訊息分析：單一品項交易訊息圖表，包含每日價量交易訊息。
4. 五大市場分析：了解單一品項於不同市場之價量關係。
5. 多品項交易訊息分析：觀察多品項於單一市場之價量關係。
6. 關聯度分析：計算供貨量及平均價 Pearson 相關係數，並繪製圖表以觀察品項之關聯性。
7. 供貨量與天氣特徵分析：觀察供貨量與天氣特徵關係。

4.1 農產品供貨量分析

欲得知台灣大宗農產品，本節將各農產品的年度供貨量及六年總供貨量，計算其總和、平均（平均計算方式：供貨量總和除以該品項交易天數，其意義為該品項平均每個交易日有多少交易量）與比例，但由於全部農產品高達 2156

項，就算將其統計結果製成圖表也難以閱讀，因此本節挑選六年總供貨量排名前 30 名農產品繪圖。

4.1.1 農產品供貨量前 30 名分析

前六年總供貨量 30 名農產品品項，結果見圖 4.2。

| | Item_NameID | Total_SUM | Total_RAT | |
|----|---------------|-------------|-----------|-------------------------------------|
| 0 | 甘藍-初秋 LA1 | 919688119.9 | 0.086592 | 15 花胡瓜 FD1 147021204.0 0.013843 |
| 1 | 包心白菜-包白 LC1 | 372132760.0 | 0.035038 | 16 胡瓜-黑刺 FC1 124967564.0 0.011766 |
| 2 | 番石榴-珍珠芭 P1 | 299159090.1 | 0.028167 | 17 花椰菜-青梗 FB1 124268006.9 0.011700 |
| 3 | 鳳梨-金鑽鳳梨 B2 | 275985004.4 | 0.025985 | 18 甜橙-柳橙 E1 120754641.1 0.011370 |
| 4 | 西瓜-大西瓜 T1 | 273707113.8 | 0.025771 | 19 蓮霧-紅蓮霧 Q1 109570872.0 0.010317 |
| 5 | 木瓜-網室紅肉 I1 | 209075327.4 | 0.019685 | 20 青江白菜-小梗 LD1 105840967.9 0.009965 |
| 6 | 蘿蔔-秆仔 SA32 | 192818838.0 | 0.018155 | 21 蕃茄-牛蕃茄 FJ3 103920971.1 0.009785 |
| 7 | 香蕉 A1 | 186722912.1 | 0.017581 | 22 苦瓜-白大米 FG1 102067598.6 0.009610 |
| 8 | 絲瓜-絲瓜 FF1 | 181478134.7 | 0.017087 | 23 玉米-甜軟殼 FY6 95582829.8 0.009000 |
| 9 | 胡蘿蔔-清洗 SB2 | 167435120.5 | 0.015765 | 24 青蔥-粉蔥 SE6 86948920.5 0.008187 |
| 10 | 甘藍-改良種 LA2 | 165195218.5 | 0.015554 | 25 青花苔 FR1 85450146.6 0.008045 |
| 11 | 其他蔬菜 OX1 | 159243568.2 | 0.014993 | 26 蘿蔔-小葉 LF2 84320301.9 0.007939 |
| 12 | 椪柑 C1 | 157533084.3 | 0.014832 | 27 芒果-愛文 R1 84059959.6 0.007915 |
| 13 | 西瓜-黃肉 T5 | 153743534.0 | 0.014476 | 28 冬瓜-白皮 FE1 82069587.8 0.007727 |
| 14 | 蘋果-富士(進口) X69 | 148778359.5 | 0.014008 | 29 甘薯-改良種 SO1 80616668.0 0.007590 |

圖 4.2 農產品供貨量前 30 名統計

統計完將結果繪製圖 4.3、圖 4.4：

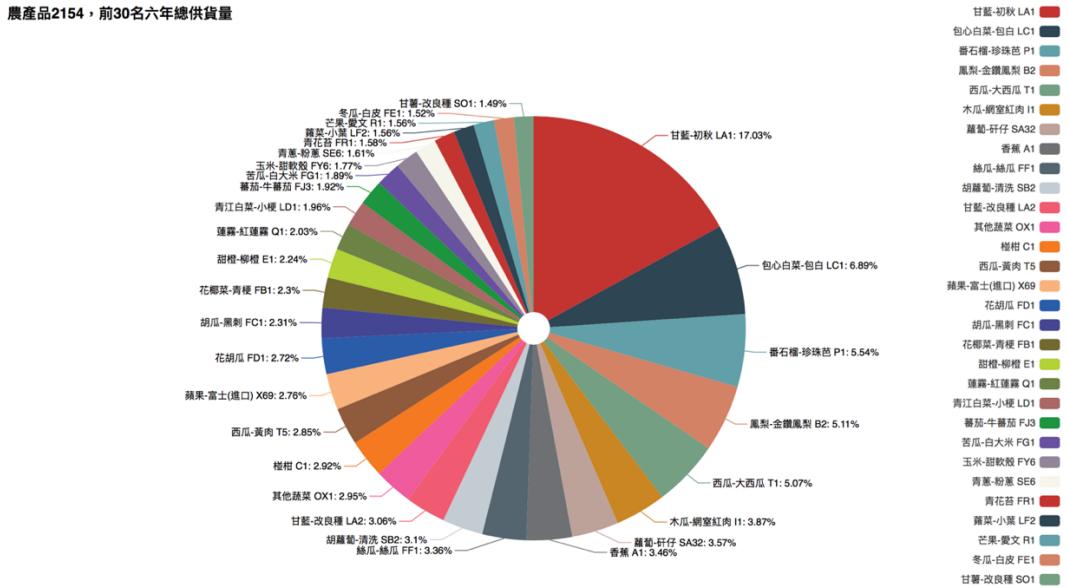


圖 4.3 農產品供貨量前 30 名圓餅圖

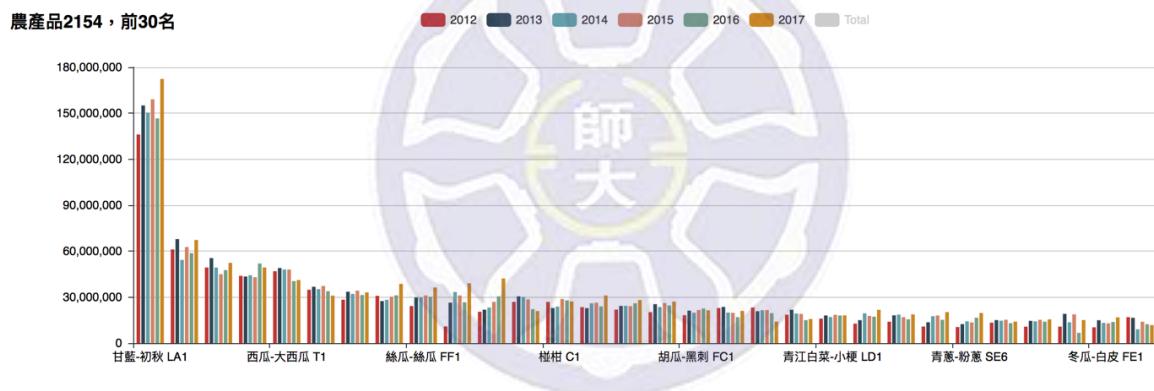


圖 4.4 農產品供貨量前 30 名長條圖

觀察圖 4.3、圖 4.4：

- 最大宗的前兩名是蔬菜：甘藍-初秋 LA1，同為歷年第一名；包心白菜 - 包白 LC1 (大白菜)，同為歷年的第 2 名。
- 第 3 到 6 名為水果：番石榴-珍珠芭 P1、鳳梨-金鑽鳳梨 B2、西瓜-大西瓜 T1、木瓜-網室紅肉 II。
- 前 30 名唯一進口農產品：蘋果-富士 (進口) X69。

4. 「胡蘿蔔-清洗 SB2」於 2012 年為第 32 名且供貨量異常的低，隔年 2013 年供貨量暴增 2.4 倍躍升至第 11 名（圖 4.5），於 2014 年供貨量持續攀升至第 7 名，至 2017 年皆為 10 名上下的農產品；同時發現另一品項「胡蘿蔔-清洗 SB21」於 2013 年為第 29 名，供貨量高於胡蘿蔔-清洗 SB2，但此後的 5 年皆不再有交易紀錄，此現象推斷可能原因：品項代號中途更換，或是該品項於 2013 年之後才確實記錄於交易平台上，即 2013 年以前的交易尚未被記錄（其他品項有此供貨量暴增的現象都有這個可能）。

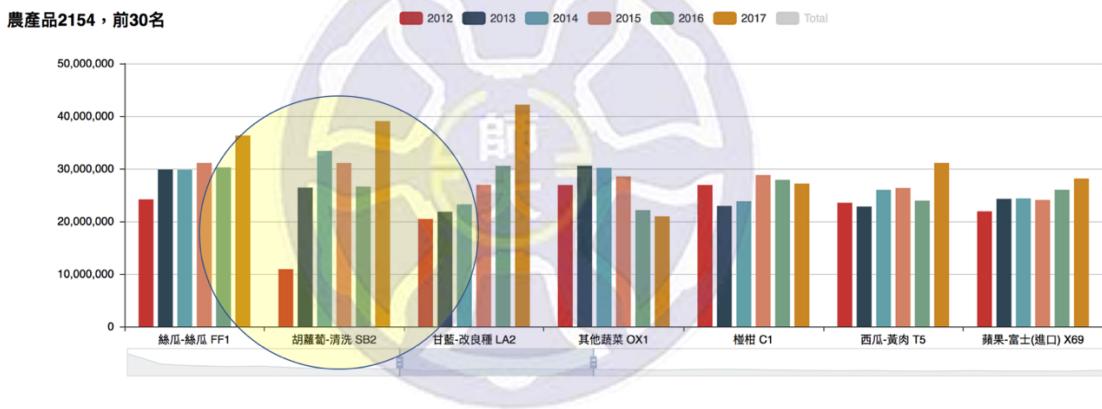


圖 4.5 胡蘿蔔-清洗 SB2 供貨量長條圖

4.1.2 蔬菜供貨量前 30 名分析

本小節先從全部品項（2156 項含蔬菜、水果及花卉）篩選出蔬菜品項（386 項）再計算供貨量總合後，挑選供貨量前 30 名蔬菜繪製圖 4.6、圖 4.7。

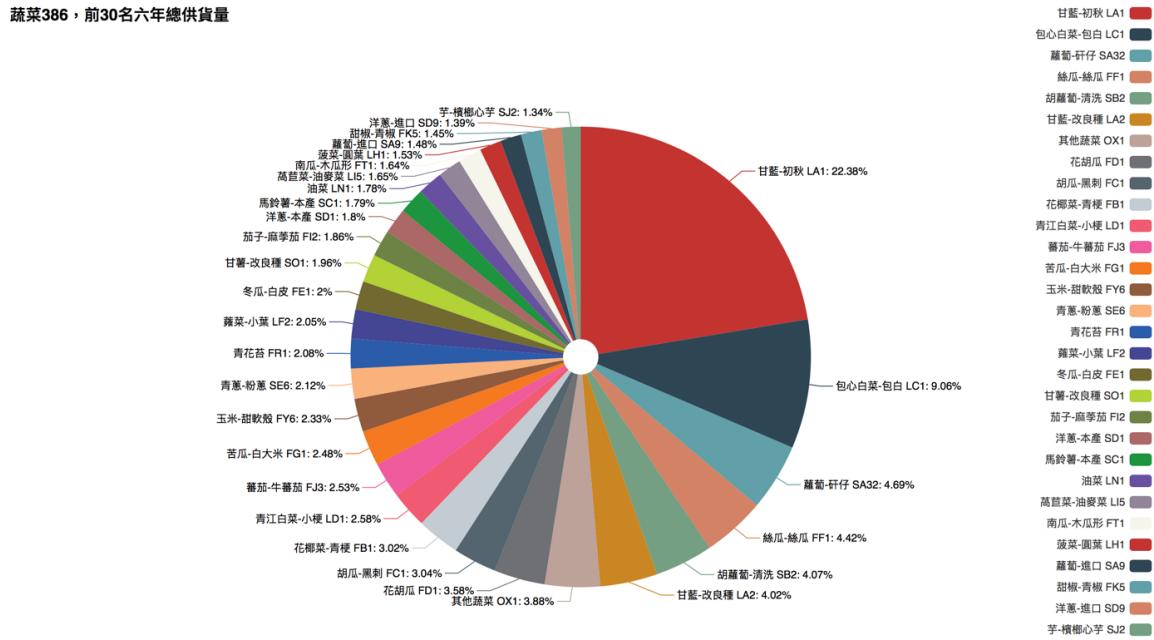


圖 4.6 蔬菜供貨量前 30 名圓餅圖

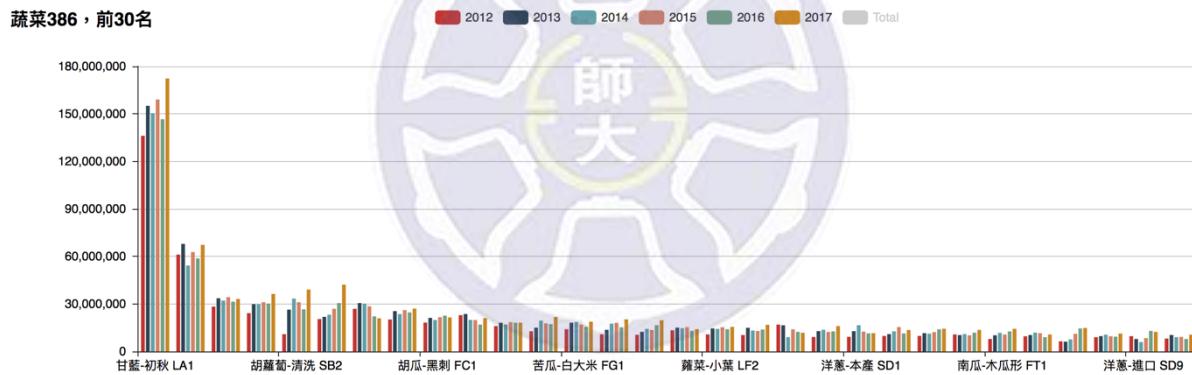


圖 4.7 蔬菜供貨量前 30 名長條圖

觀察圖 4.6、圖 4.7：

1. 供貨量前 7 名佔超過前 30 名的一半：甘藍-初秋 LA1、包心白菜-包白 LC1、蘿蔔-研仔 SA32 (白蘿蔔)、絲瓜-絲瓜 FF1、胡蘿蔔-清洗 SB2、其他蔬菜 OX1。
2. 前 30 名蔬菜進口包括：第 27 名「蘿蔔-進口 SA9」；第 30 名「洋蔥-進

口 SD9」。

3. 第6名：甘藍-改良種 LA2 同為甘藍但品種不同，且供貨量逐年穩定提高（圖4.8、圖4.9）。



圖 4.8 甘藍-改良種 LA2 供貨量長條圖 1



圖 4.9 甘藍-改良種 LA2 供貨量長條圖 2

4.2 市場供貨量分析

欲了解各市場交易比例，本節將各個市場的年度供貨量及六年總供貨量，計算其總和、平均（總和除以該品項於該年不含市場休息的天數，其意義為該市場平均每個交易日的交易量）。

4.2.1 全部市場供貨量分析

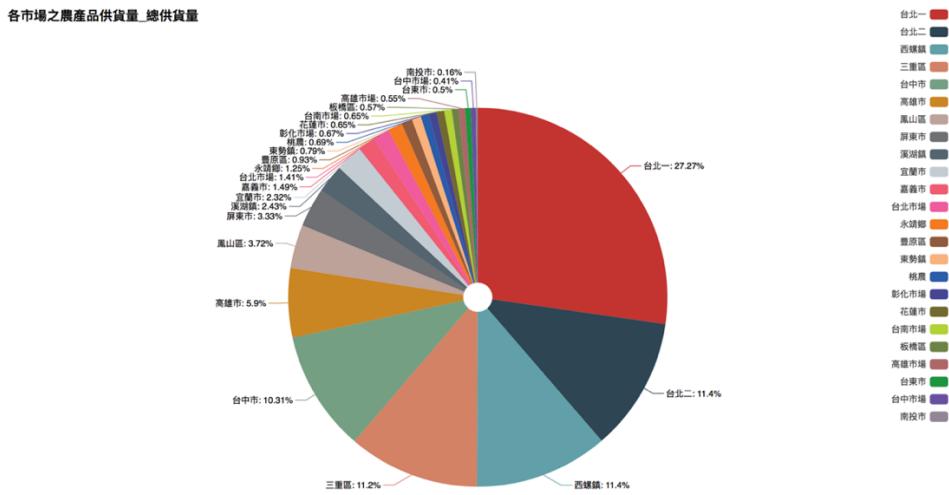


圖 4.10 全部市場供貨量圓餅圖

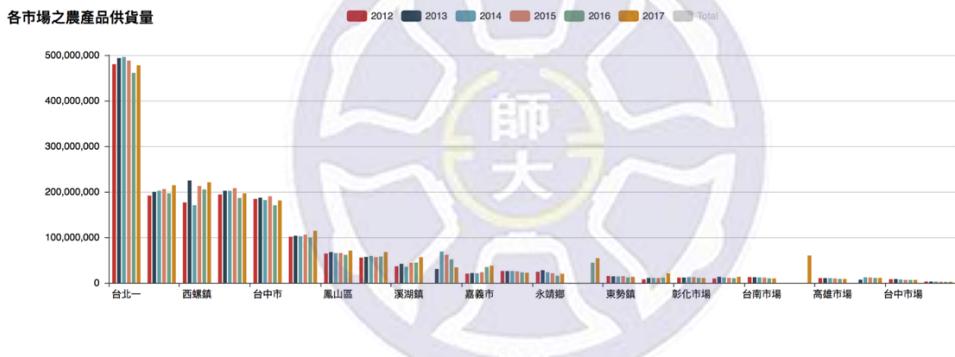


圖 4.11 全部市場供貨量長條圖

觀察圖 4.10、圖 4.11 可知前 3 名市場分別為台北一、台北二、西螺鎮，佔總和為 50.4%，即台北兩大市場和雲林西螺鎮市場之農產品交易量佔全台灣一半以上。

4.2.2 蔬菜市場分析

本研究主要針對國產蔬菜價格預測，本小節篩選出蔬菜品項（386項）再進行加總與繪圖（圖 4.12、圖 4.13）。

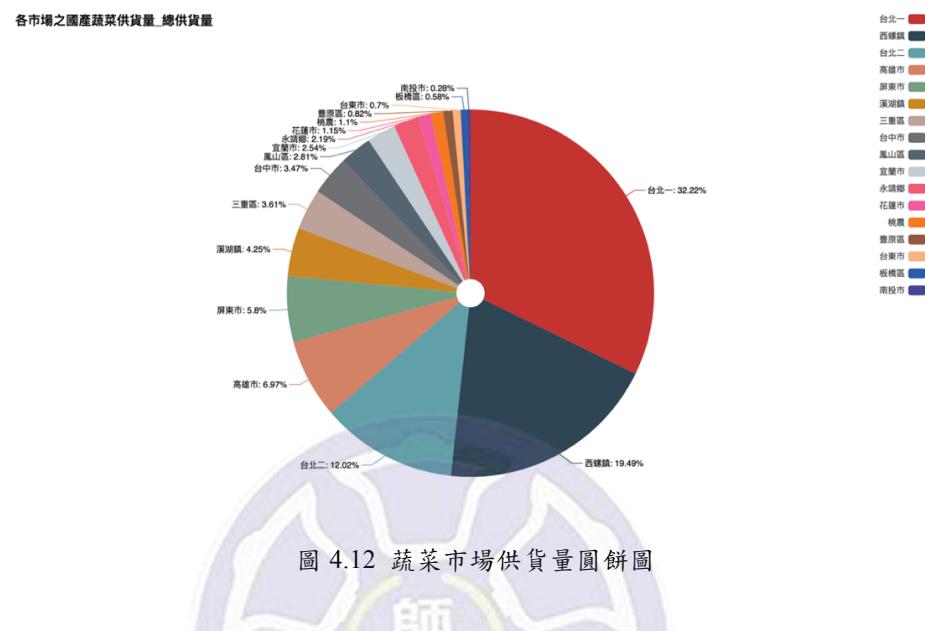


圖 4.12 蔬菜市場供貨量圓餅圖

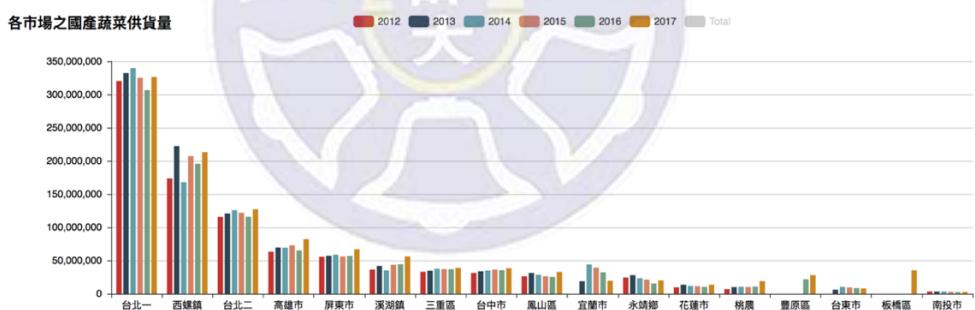


圖 4.13 蔬菜市場供貨量長條圖

觀察圖 4.12、圖 4.13 發現全部市場中有交易國產蔬菜台灣共 17 個市場；前 3 名分別為台北一、西螺鎮、台北二，西螺鎮原先為第 3 名，可見西螺鎮國產蔬菜供貨量比例高於水果及花卉，亦可解釋為台北二的水果和花卉的供貨量較多才能在全部品項中佔第 2 名。

4.3 單一品項交易訊息分析

欲了解單一品項於單一市場之交易狀態，本小節針對單一品項繪製三種圖表：各年總供貨量比例、歷年供貨量及平均價、每日交易資訊分析。品項挑選供貨量最高同時為本研究目標的「甘藍-初秋 LA1」，針對全台供貨量最大的「台北一」市場。

4.3.1 各年總供貨量比例

統計單一品項（甘藍-初秋 LA1）於單一市場（台北一）之各年度總供貨量，觀察圖 4.14 可見甘藍-初秋 LA1 年度總供貨量十分穩定。

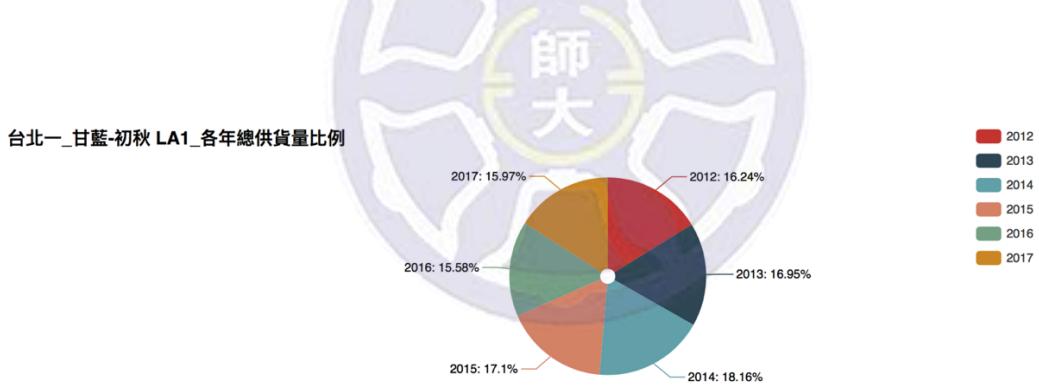


圖 4.14 甘藍-初秋 LA1 歷年總供貨量圓餅圖

4.3.2 歷年供貨量及平均價

為了觀察甘藍-初秋 LA1 供貨量與平均價的趨勢，分別繪製成圖 4.15。橫軸為日期；縱軸為供貨量（公斤）/價格（元）；不同顏色代表不同年份。本圖表

之歷年供貨量為堆積圖，6 年堆疊成一柱；也可以切換成平坦圖（圖 4.16），可以隨著觀察方式自由切換。

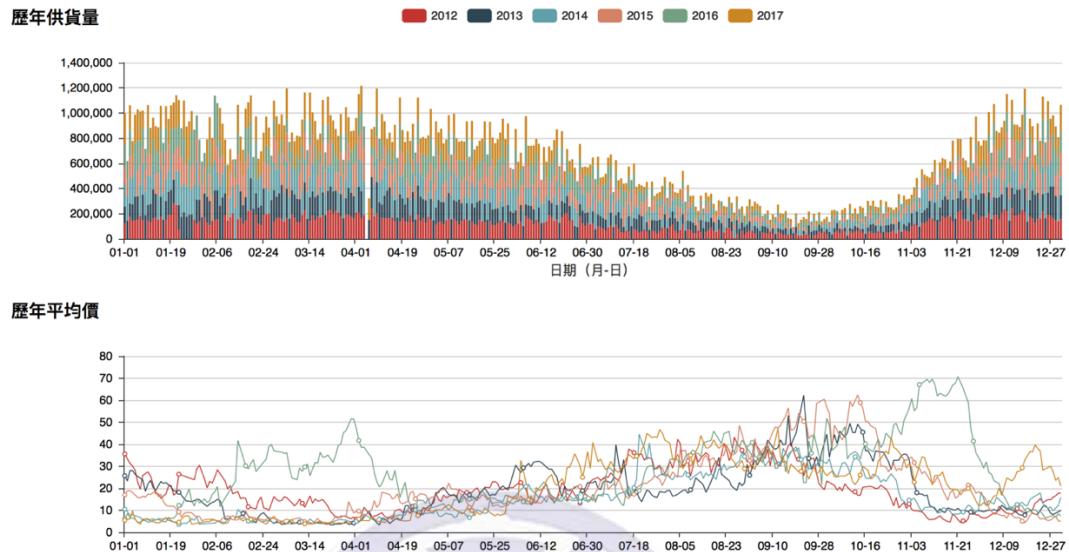


圖 4.15 甘藍-初秋 LA1 歷年供貨量（堆積圖）及平均價

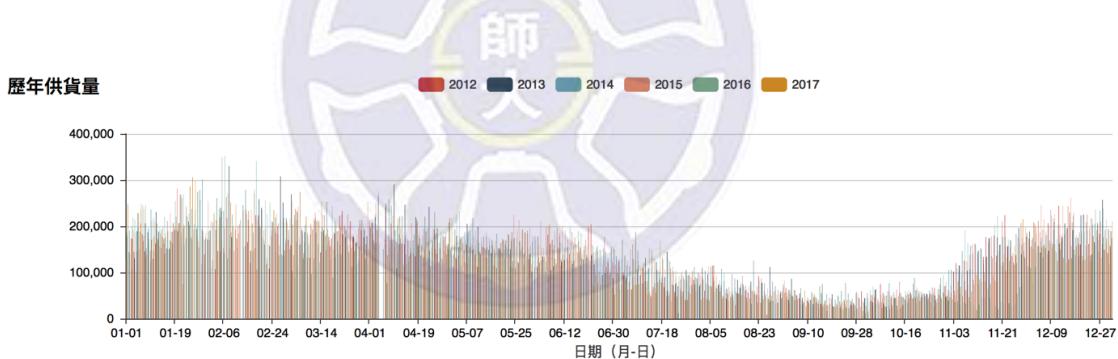


圖 4.16 甘藍-初秋 LA1 歷年供貨量（平坦圖）

由兩種類型的歷年供貨量圖表可見供貨量呈現週期性波動，顯示甘藍-初秋 LA1 具季節性，供貨量於每年 6 月開始遞減至 10 月，隨後往上攀升。

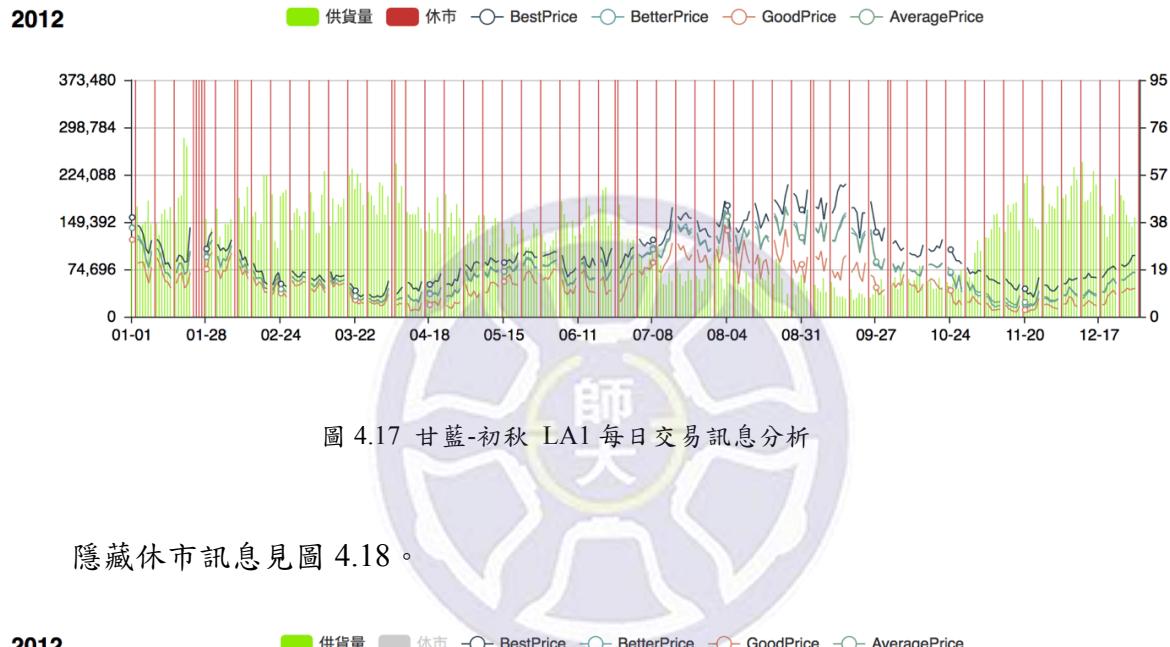
4.3.3 每日交易訊息分析

將供貨量與 4 種價格（上、中、下、平均價），及「休市特徵」合併繪製

圖 4.17。橫軸為日期，綠色長條圖為供貨量，對應左側縱軸、紅線為市場休息

日，折線圖為四種價格，對應右側縱軸，圖形上方的標籤皆可點選並自由切換

觀察目標。



隱藏休市訊息見圖 4.18。

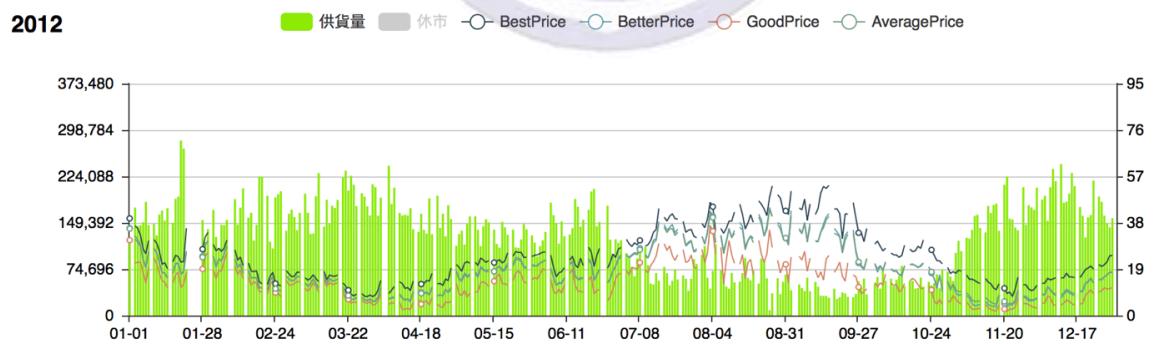


圖 4.18 甘藍-初秋 LA1 每日交易訊息分析-隱藏休市訊息

觀察發現每年甘藍-初秋 LA1 的淡季供貨量短少，價格也會隨著提高，符合

市場機制：供給少，價格高。

4.4 五大市場分析

欲了解單一品項於其他市場的交易狀態，由「4.2.2 蔬菜市場分析」可知全台灣前五大市場為：台北一（32.22%）、西螺鎮（19.49%）、台北二（12.02%）、高雄市（6.97%）、屏東市（5.8%），此五大市場佔國產蔬菜市場76.49%，因此本研究於多市場觀察中僅觀察此5大市場。本小節挑選甘藍-初秋LA1為觀察對象，並繪製五大市場各年供貨量比例，見圖4.19。

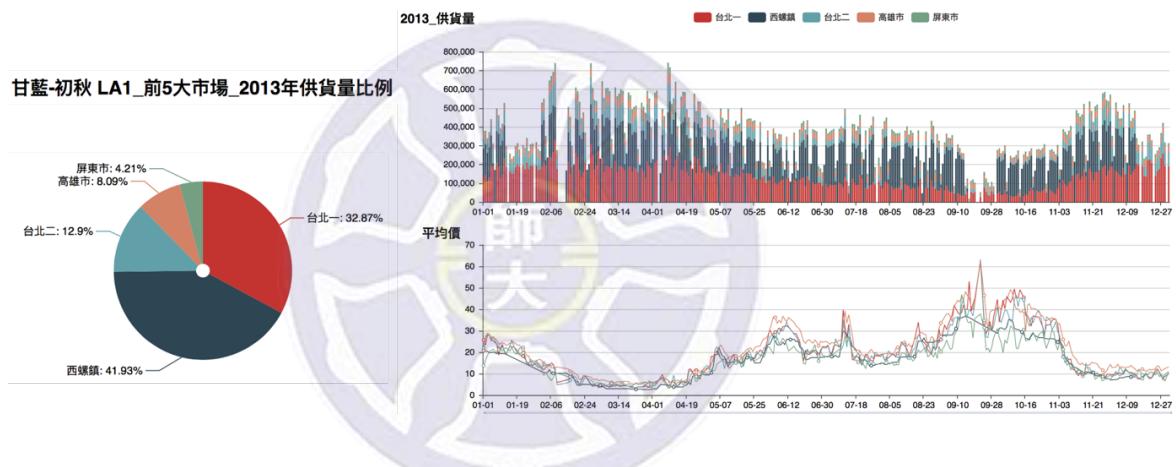


圖 4.19 甘藍-初秋 LA1 五大市場分析

觀察發現西螺鎮的供貨量大於台北一，但西螺鎮人口不大能夠消化這麼大量的農產品，由此假設西螺鎮可能為台北一市場的上游，甚至是台灣各地果菜市場的源頭。

4.5 多品項交易訊息分析

由於不同種類農產品或是同種但不同品項間存在差異差異，為了觀察不同品項之交易價量關係，於是將不同品項合併繪圖。本小節以胡蘿蔔為例：胡蘿蔔-清洗 SB2、胡蘿蔔-未洗 SB1、胡蘿蔔-未洗 SB11、胡蘿蔔-清洗 SB21（圖 4.20）。

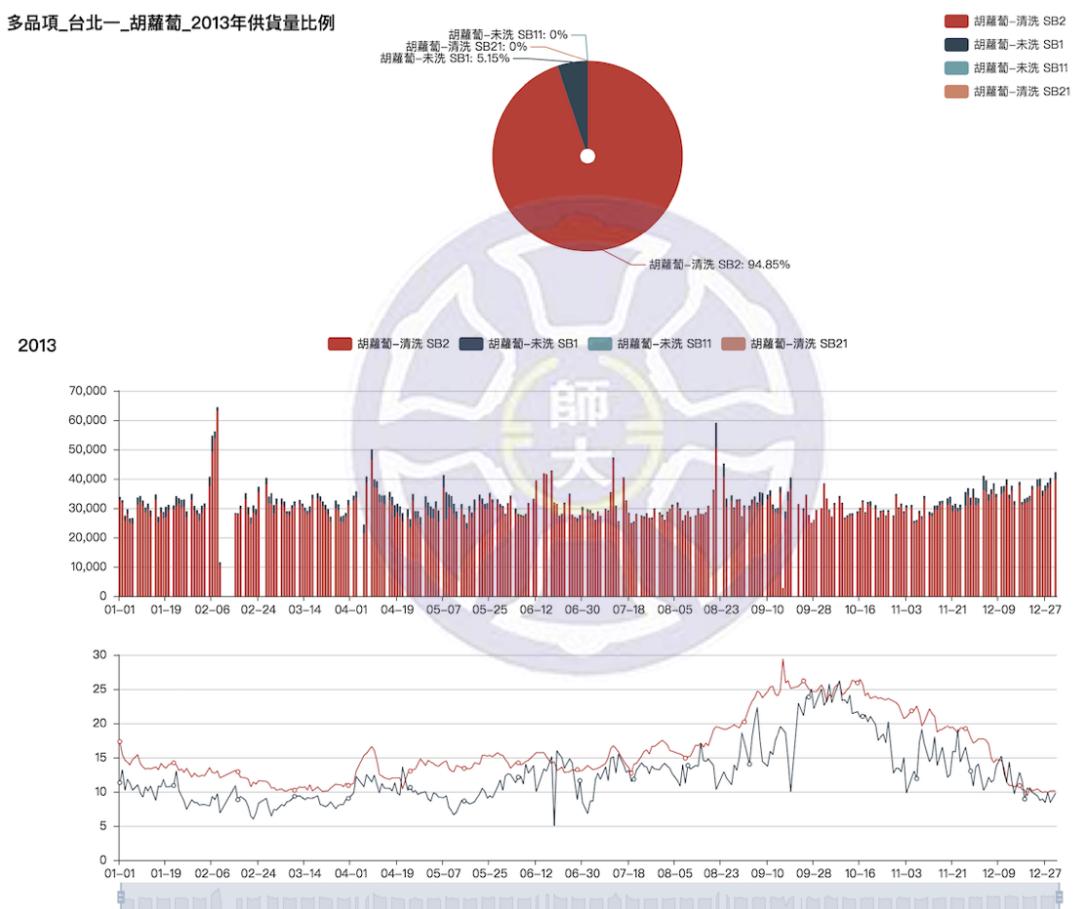


圖 4.20 胡蘿蔔之多品項交易訊息分析

4.6 關聯度分析

欲瞭解單一市場內之單一品項與其他品項之供貨量及價格關係，本小節針對台北一市場之國產蔬菜（349項）進行供貨量與平均價關聯度分析，資料操作步驟如下：

步驟一：挑選出台北一市場農產品。

步驟二：挑選出國產蔬菜349項。

步驟三：分別挑出變量：供貨量/平均價數據。

步驟四：計算兩兩品項變量之 Pearson 相關係數。

步驟五：繪製視覺化圖表。

步驟六：結合「4.5 多品項交易訊息分析」，進一步觀察相關度高/低的多品項分析。

Pearson 相關係數（Pearson product-moment correlation coefficient）用於分析兩變異數之間的線性相關性。其值介於1至-1之間。此處的相關性不代表兩變量之間互相具影響性，僅能得知一變量提高，另一品項也提高，或同時下降。

4.6.1 國產蔬菜供貨量關聯度分析

計算完 Pearson 相關係數後繪製圖 4.21，兩座標軸表示各品項；內部顏色代表兩兩品項之關聯度，Pearson 相關係數越接近1顏色越紅；越接近-1顏色越接近藍綠色。

台北一：大宗蔬菜_Amount_相關係數

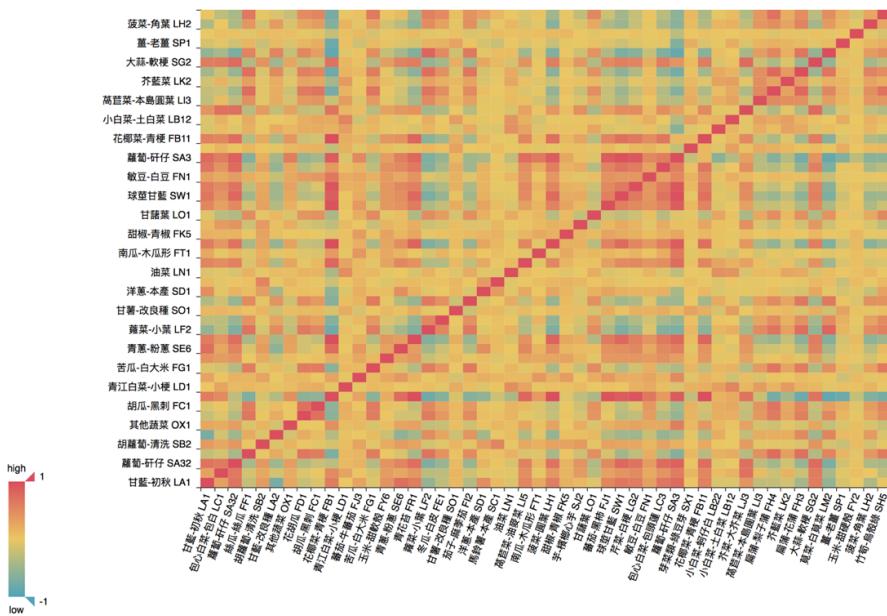


圖 4.21 國產蔬菜供貨量關聯度分析

此圖表可針對關聯度設定範圍，如選取低相關度（圖 4.22）；亦可縮放圖

片觀察目標品項（圖 4.23）。

台北一：大宗蔬菜_Amount_相關係數

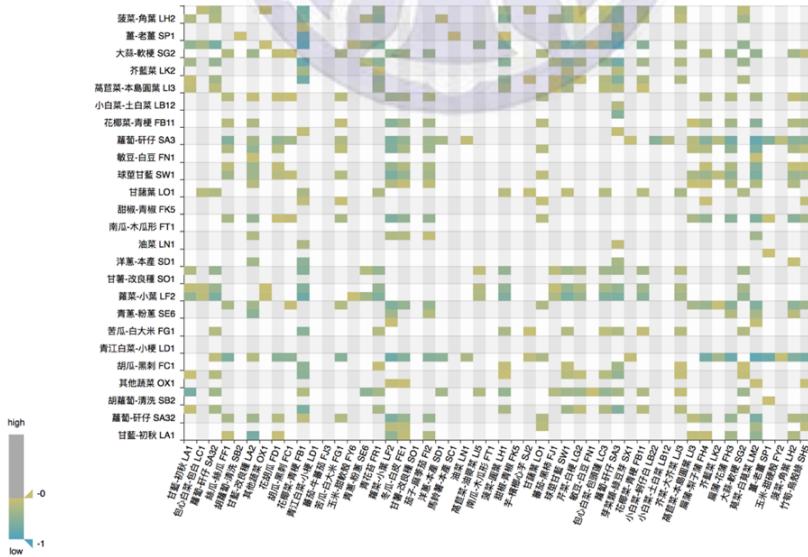


圖 4.22 國產蔬菜供貨量關聯度分析-篩選低關聯度

台北一：大宗蔬菜_Amount_相關係數

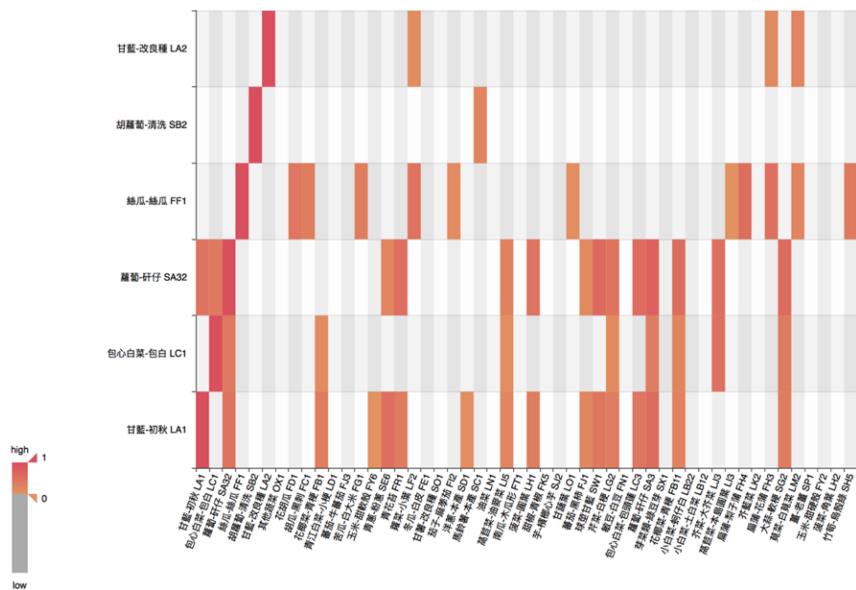


圖 4.23 國產蔬菜供貨量關聯度分析-放大圖

4.6.2 國產蔬菜平均價關聯度分析

台北一：大宗蔬菜_AVEPrice_相關係數

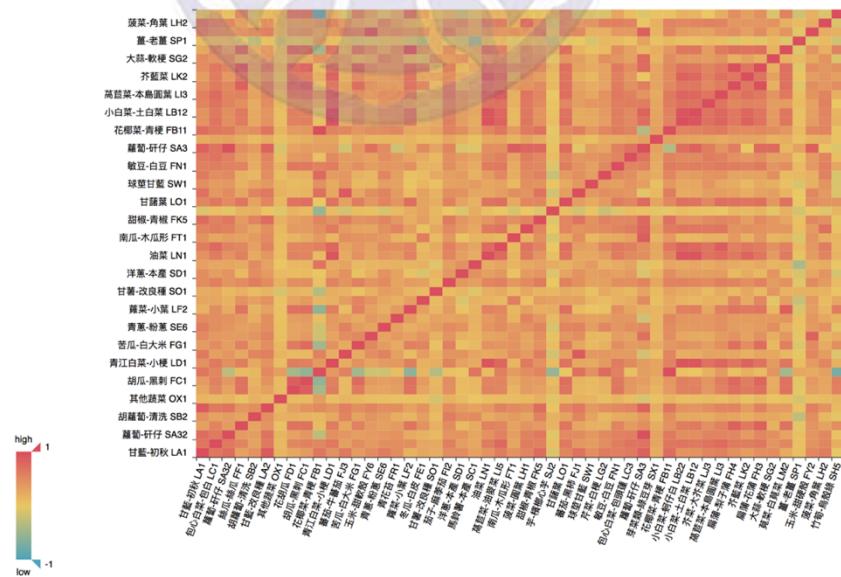


圖 4.24 國產蔬菜平均價關聯度分析

由平均價關聯度分析（圖 4.24）發現平均價的色溫高於供貨量關聯度（圖 4.21），代表菜價的相依性普遍較高。

4.6.3 進階多品項分析

針對目標甘藍-初秋 LA1，挑選其相關度高或低之蔬菜品項，再結合「4.5 多品項交易訊息分析」進一步觀察這些具高/低相關的品項。

1. 甘藍-初秋 LA1 供貨量高相關前 5 名品項（圖 4.25）：蘿蔔-研仔

SA32、青蔥-粉蔥 SE6、球莖甘藍 SW1、包心白菜-包頭蓮 LC3、蘿蔔

-研仔 SA3，再繪製多品項交易訊息分析（圖 4.26），由於甘藍-初秋

LA1 與其他品項供貨量差距甚大，因此本小節皆隱藏甘藍-初秋 LA1，

以利於觀察。

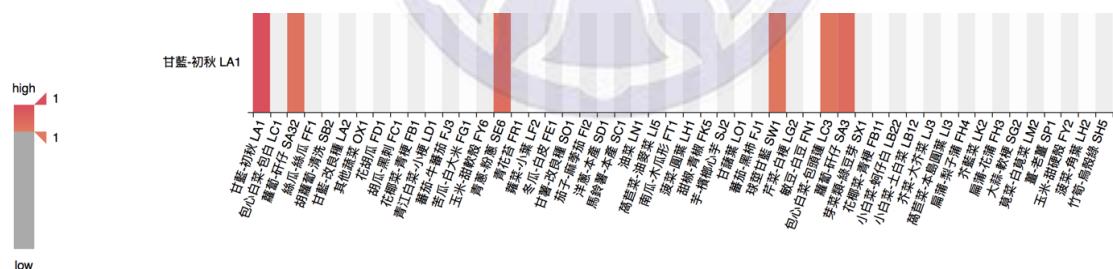


圖 4.25 甘藍-初秋 LA1 供貨量高相關前 5 名品項

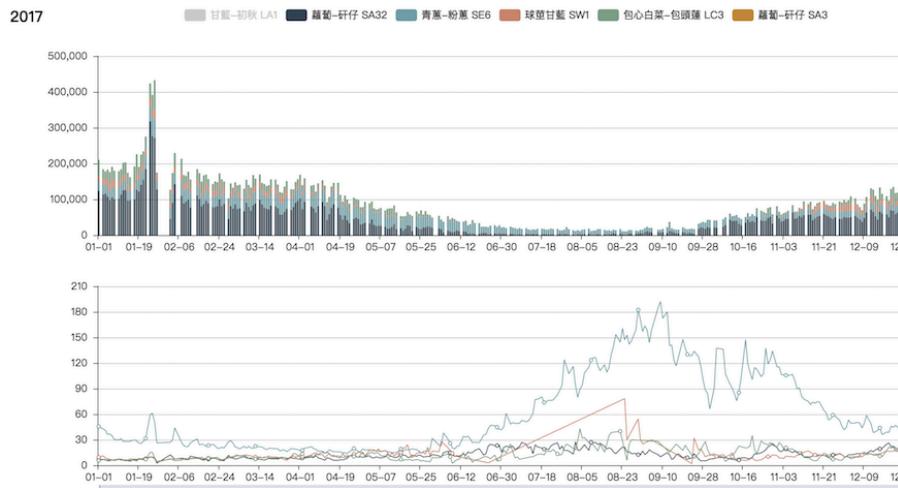


圖 4.26 甘藍-初秋 LA1 供貨量高相關前 5 名-多品項交易訊息分析

2. 甘藍-初秋 LA1 供貨量低相關前 5 名品項圖 4.27：甘藍-改良種 LA2、

蘿菜-小葉 LF2、扁蒲-花蒲 FH3、莧菜-白莧菜 LM2、菠菜-角葉

LH2。觀察圖 4.28 發現部分品項平均價之漲跌幅度甚大，不同於甘藍-

初秋 LA1 的價格具整年的高低週期，如「菠菜-角葉 LH2」。放大細看

後發現可能因為該供貨量起伏大，連帶造成價格波動。這些價量波動

過大的品項可能不利於納入目標價格預測特徵，原因是其價格成因不

確定因素過多，但個別來看仍有期研究價值。

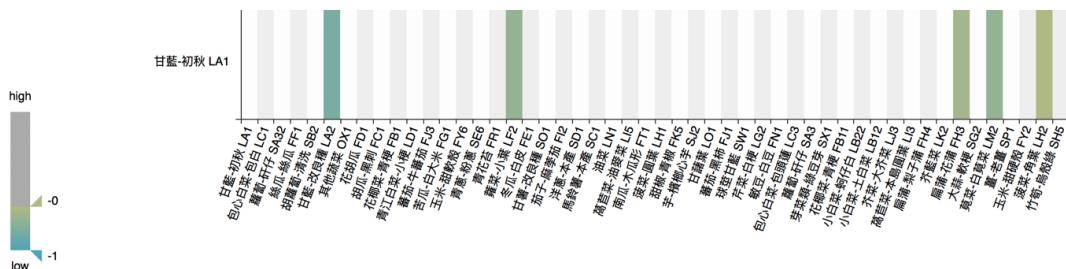


圖 4.27 甘藍-初秋 LA1 供貨量低相關前 5 名品項

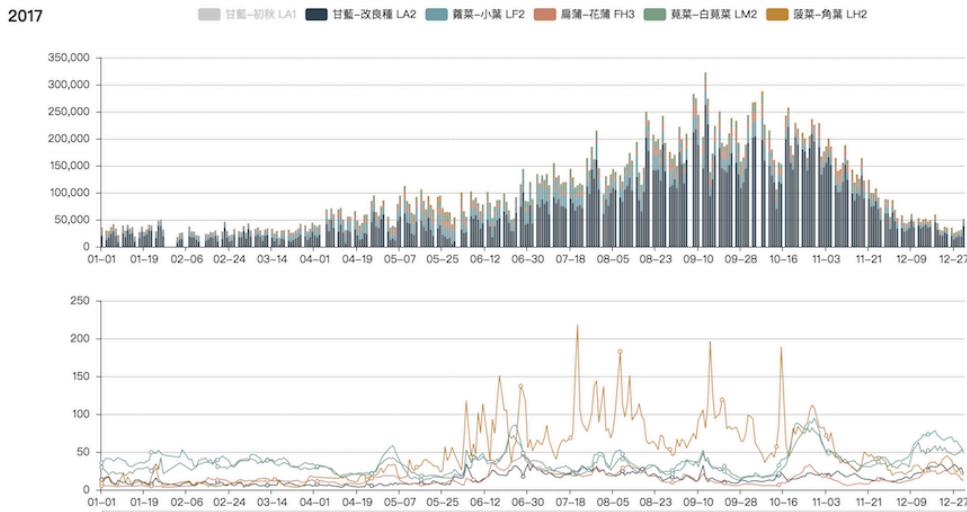


圖 4.28 甘藍-初秋 LA1 供貨量低相關前 5 名-多品項交易訊息分析

3. 甘藍-初秋 LA1 平均價高相關前 5 名品項（圖 4.29）：蘿蔔-研仔

SA32、甘藍-改良種 LA2、甜椒-青椒 FK5、蘿蔔-研仔 SA3、花椰菜-青梗 FB11。綜合供貨量低相關，觀察發現「甘藍-改良種 LA2」同時為甘藍-初秋 LA1「供貨量低相關」與「平均價高相關」。表示甘藍-改良種 LA2 價格波動非供應少而價格高，而是由於 6 月中到 11 月為甘藍-初秋 LA1 的淡季，使得甘藍改良種相對需求提高，儘管為本身旺季

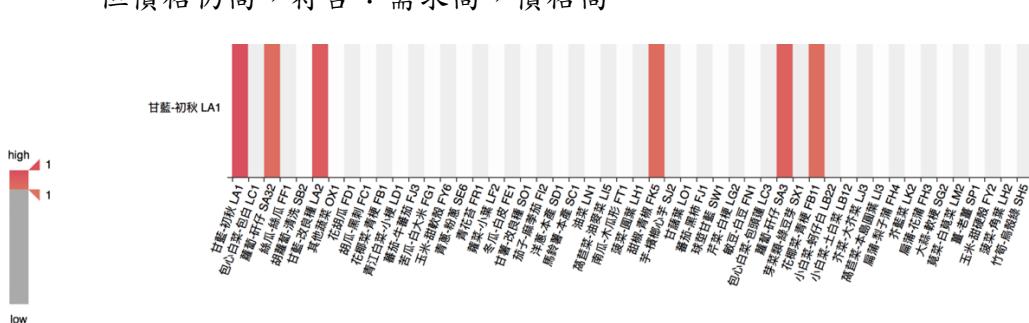


圖 4.29 甘藍-初秋 LA1 平均價高相關前 5 名品項

2017 甘藍-初秋 LA1 蘿蔴-小葉 LF2 甘薯-改良種 SO1 芋-檳榔心芋 SJ2 薑-老薑 SP1 竹筍-烏殼綠 SH5

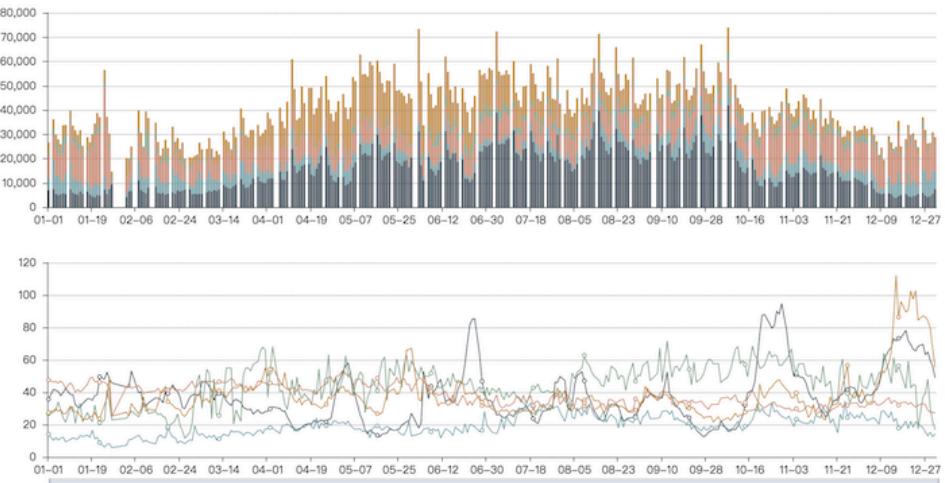


圖 4.30 甘藍-初秋 LA1 平均價高相關前 5 名-多品項交易訊息分析

4. 甘藍-初秋 LA1 平均價低相關前 5 名品項 (圖 4.31)：蘿蔴-小葉

LF2、甘薯-改良種 SO1、芋-檳榔心芋 SJ2、薑-老薑 SP1、竹筍-烏殼綠 SH5，由於不清楚其他蔬菜為何，因此「其他蔬菜 OX1 」不納入考量。

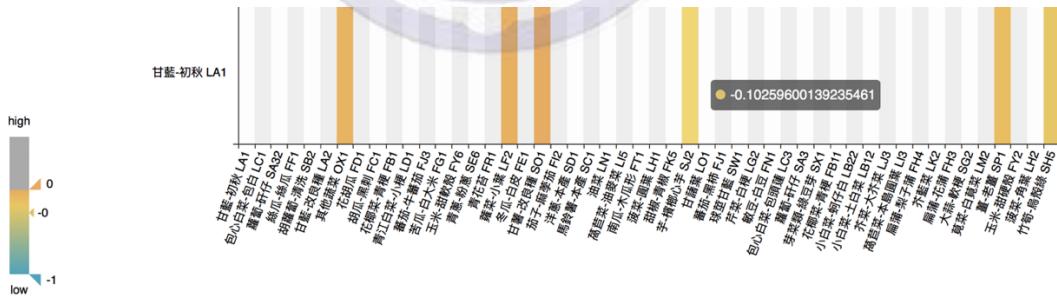


圖 4.31 甘藍-初秋 LA1 平均價低相關前 5 名品項

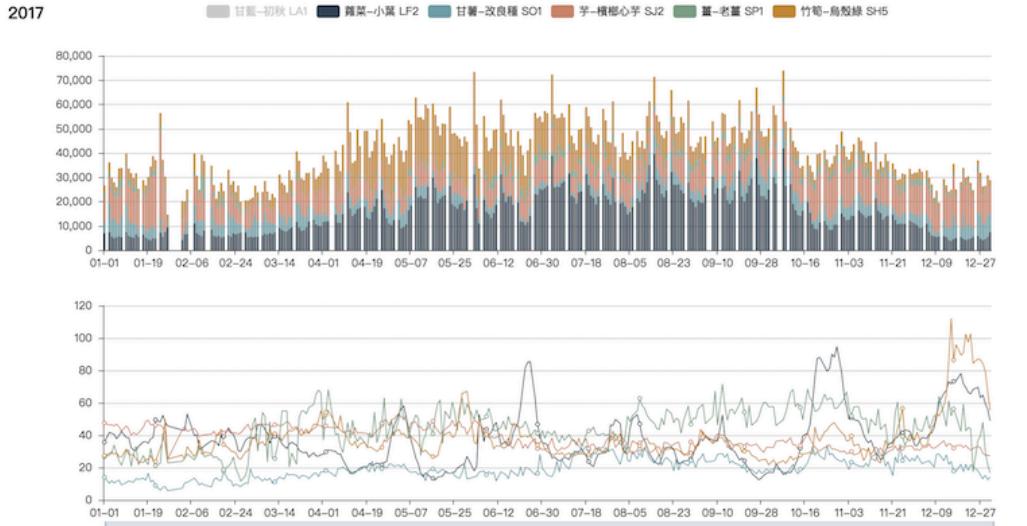


圖 4.32 甘藍-初秋 LA1 平均價低相關前 5 名-多品項交易訊息分析

本小節於篩選品項時發現甘藍-初秋 LA1 平均價低相關度僅有 2 品項為負值 (Pearson 相關係數)，其他均為正，可解讀為國產蔬菜平均價泛正相依，即大部分的菜價波動較為相似。

4.7 供貨量與天氣特徵分析

由於植物生長與環境變化息息相關，因此將農產品交易開放資料結合氣象開放資料進行視覺化呈現，基於植物生長需要過去天氣現象的累積，因此將個別天氣特徵的過去數值分別加總與平均，以模擬其生長的環境，本研究在此按照過去 15、30、45、60 天進行統計，為此額外擷取將 2011 年下半年天氣資料。另外根據自由時報「高麗菜有這麼多種？口感特色一次看懂！」[5]：

「秋冬時節，高麗菜是台灣平地常見的蔬菜，彰化、雲林、嘉義為主產地區；夏季則在宜蘭、台中、南投等海拔約 1500 公尺的高冷地栽培，價格較高一些，由於夏季種植面積較少，因此部分需仰賴進口。」

因此本研究挑選彰化縣、雲林縣、嘉義縣、宜蘭縣、台中市、南投縣作為視覺化目標地區，並合併颱風警報特徵繪圖，綜合觀察圖表：

1. 觀察圖 4.33、圖 4.34、圖 4.37、圖 4.38 累積值與平均值差異，降水量兩種統計方式呈現兩種波形，因此降水量部分可以考慮累積值與平均值兩種特徵變換方式。
2. 承 1，觀察圖 4.35、圖 4.36、圖 4.39、圖 4.40，溫度累積值對於歷史天氣的變化幅度呈現力大於平均值，因此可以將溫度累積值納入實驗。
3. 降水量與溫度相比較，單獨觀察整年度可以發現溫度的波形與供貨量趨勢成反相關，直覺來看溫度的影響力高於降水量。



圖 4.33 彰化、雲林、嘉義累積雨量分析

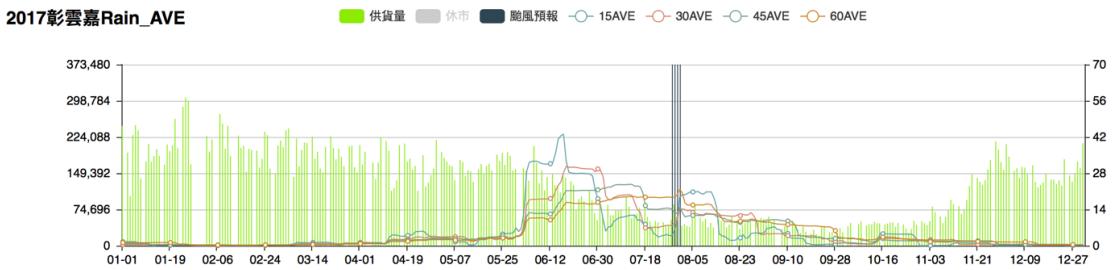


圖 4.34 彰化、雲林、嘉義平均雨量分析

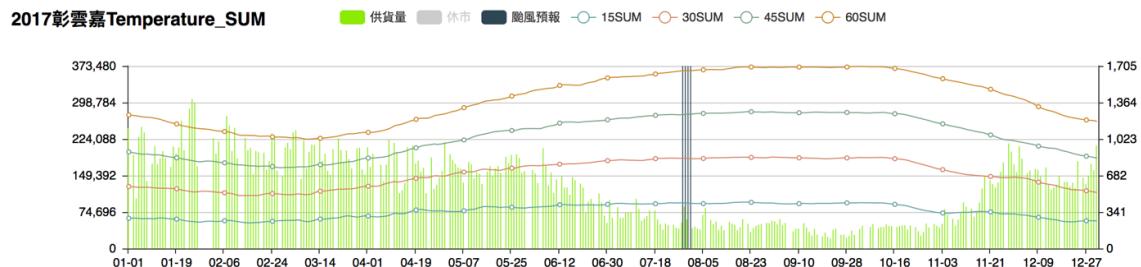


圖 4.35 彰化、雲林、嘉義累積溫度分析

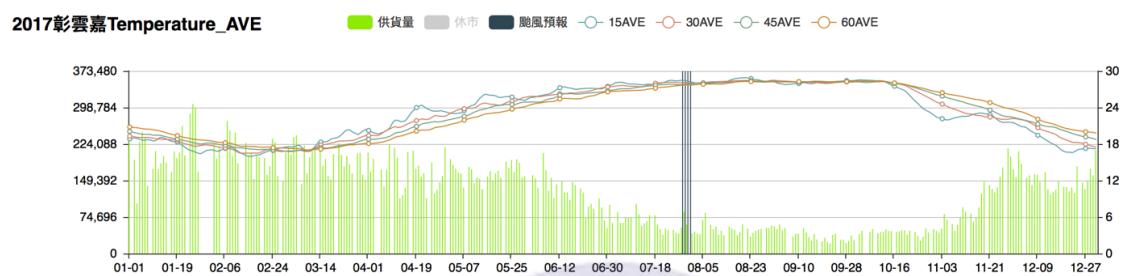


圖 4.36 彰化、雲林、嘉義平均溫度分析



圖 4.37 宜蘭、台中、南投累積雨量分析

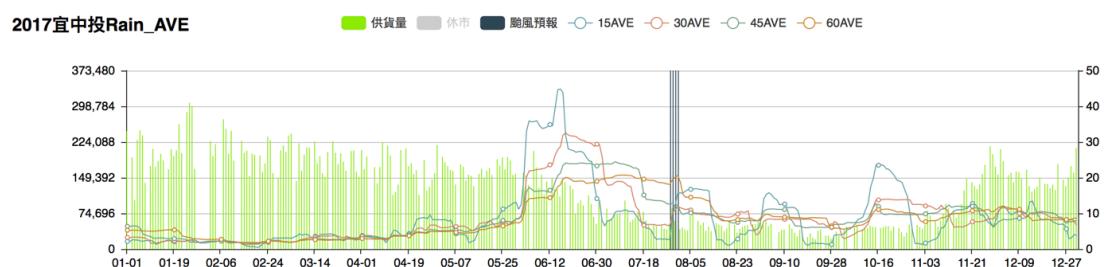


圖 4.38 宜蘭、台中、南投平均雨量分析

2017宜中投Temperature_SUM

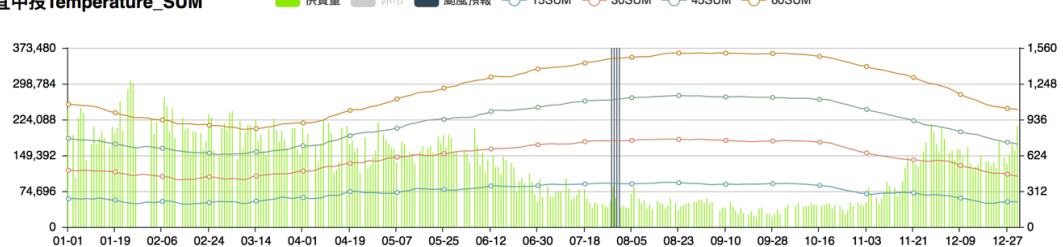


圖 4.39 宜蘭、台中、南投累積溫度分析

2017宜中投Temperature_AVE

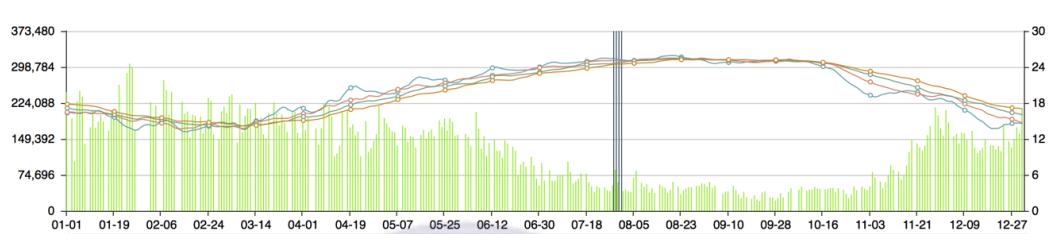


圖 4.40 宜蘭、台中、南投平均溫度分析



第五章、特徵分析

本研究目標預測台北一市場之甘藍-初秋 LA1 的平均價，經過上述開放資料擷取與特徵工程整理出特徵包含上價、中價、下價、供貨量、19 個縣市溫度、19 個縣市降水量、颱風警報發布，另外建構出市場休息特徵、日期特徵。但是為了提高價格預測模型可信度與準確度，本章模擬不同特徵於生活中影響價格的現實情境，將各種不同特徵按其特性採取不同特徵變換建立出「候選特徵」，再藉由隨機森林演算法進行價格漲跌分類實驗，並觀察候選特徵對於價格漲跌的分類效果，挑選出價格預測模型之最終特徵。

價格漲跌分類實驗採用隨機森林演算法，經過多次實驗發現於此實驗中的分類型實驗準確率高於數值型回歸預測，故將本章將實驗設計成分類型實驗。實驗流程與參數設定如下：

1. 每次實驗共 1959 筆資料（自 2012 年 1 月 1 日起，2018 年 7 月 3 日迄，原 2352 筆扣除休市日期剩下 1959 筆），實驗前將整份資料先打散（shuffle）再拆成訓練集（佔總資料量七成：1371 筆）與測試集（佔總資料量三成：587 筆）。
2. 由於本實驗目的僅在於挑選出價格分類效果好的特徵，因此每次實驗的樹數量為 100 顆。
3. 每次實驗結果有兩類：分類準確度及候選特徵分類權重，同時觀察兩結果進而挑選出最終模型採納的特徵。

5.1 平均價漲跌分類方法

本小節針對甘藍-初秋 LA1 平均價進行漲跌分類實驗，將連續型的價格資料轉換成類別型資料，本方法漲跌定義先計算當日價格在前一段時間範圍內（歷史價格）的平均與離散程度（計算方法同標準差計算方式：計算前幾日價格與平均差值之平方和的平均），在判斷當日價格與評價假的差值與離散程度之關係（由於市場休息日會中斷連續交易日期，因此本方法將忽略休市日來計算天數）。

分類方法為當日價格減去前 x 日價格之價格平均，再將差值和前 x 日價格之離散程度比較。例如，當日菜價為 30 元，前三日平均為 25 元、離散程度為 4 元，當日菜價與菜價平均差值為 +5 元（大於離散程度 4 元），表示為該日價格於前三日交易價格漲超過一個離散程度值。

經由多次測試實驗後發現漲跌分類數量過多或漲跌分類數量不平衡皆使得價格漲跌分類預測準確率下降，因此對於價格分類的參數進行討論：

1. 歷史價格範圍：價格若於週期內波動過大，使離散程度提高，進而使漲價或跌價的敏銳度下降，因此歷史價格範圍不宜過大。
2. 離散程度區間：縮小離散程度區間進行分類，可以提高價格波動的敏銳度，實際上也可以更細微地預測價格漲跌幅度，甚至減緩分類不平橫現象，但會造成實驗進行多分類預測，使得準確率驟降，因此離散程度區間不宜過小。

本方法最終針對歷史價格範圍定為前 3 日和前 10 日價格，價格漲跌分類標準見表 5.1。

表 5.1 價格漲跌分類規則

| 漲跌狀態 | 賦值 |
|-------------------|----|
| 漲超 0.7 個離散程度 | 3 |
| 價格於上下 0.7 個離散程度之間 | 2 |
| 跌超 0.7 個離散程度 | 1 |

前 3 日與前 10 日平均價漲跌三分類結果之個數關係：圖 5.1、圖 5.2。（長條圖表示該分類總個數）

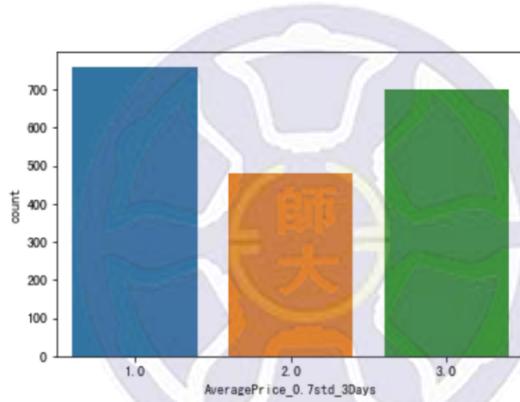


圖 5.1 平均價漲跌三分類-前 3 日價格

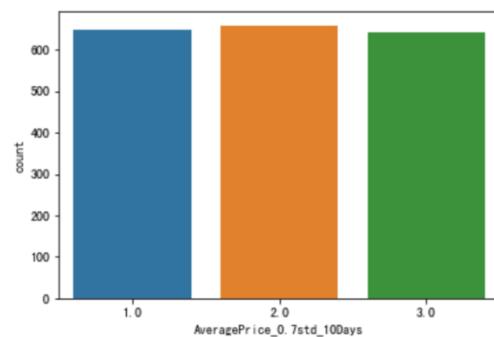


圖 5.2 平均價漲跌三分類-前 10 日價格

5.2 供貨量特徵

農產品價格影響最主要原因為當日供貨量多寡，本小節假設近期價格亦受前一段時間的累積供貨量影響。因此將 1 至 20 日之連續時間內的供貨量加總，再進行累積供貨量之平均價漲跌分類實驗與驗證，結果見表 5.2。

表 5.2 累積供貨量實驗結果

| 累積供貨量 | | 累積供貨量 | |
|-------------------------|--------------|-------------------------|--------------|
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| Amount_1Days_CNT | 0.133 | Amount_1Days_CNT | 0.093 |
| Amount_2Days_CNT | 0.072 | Amount_2Days_CNT | 0.071 |
| Amount_3Days_CNT | 0.057 | Amount_4Days_CNT | 0.058 |
| Amount_5Days_CNT | 0.051 | Amount_3Days_CNT | 0.056 |
| Amount_4Days_CNT | 0.049 | Amount_20Days_CNT | 0.054 |
| Amount_8Days_CNT | 0.048 | Amount_19Days_CNT | 0.053 |
| Amount_6Days_CNT | 0.047 | Amount_5Days_CNT | 0.051 |
| Amount_7Days_CNT | 0.045 | Amount_6Days_CNT | 0.048 |
| Amount_20Days_CNT | 0.044 | Amount_18Days_CNT | 0.048 |
| Amount_9Days_CNT | 0.043 | Amount_17Days_CNT | 0.046 |
| Amount_10Days_CNT | 0.043 | Amount_16Days_CNT | 0.044 |
| Amount_12Days_CNT | 0.042 | Amount_7Days_CNT | 0.043 |
| Amount_13Days_CNT | 0.041 | Amount_8Days_CNT | 0.043 |
| Amount_14Days_CNT | 0.040 | Amount_15Days_CNT | 0.043 |
| Amount_19Days_CNT | 0.040 | Amount_12Days_CNT | 0.042 |
| Amount_17Days_CNT | 0.040 | Amount_13Days_CNT | 0.041 |
| Amount_18Days_CNT | 0.040 | Amount_10Days_CNT | 0.039 |
| Amount_15Days_CNT | 0.040 | Amount_9Days_CNT | 0.039 |
| Amount_11Days_CNT | 0.039 | Amount_14Days_CNT | 0.039 |
| Amount_16Days_CNT | 0.035 | Amount_11Days_CNT | 0.039 |

實驗結果可看出當日供貨量對於漲跌分類權重最高，其次為近五日內累積供貨量；於「10 日漲跌分類」實驗中，累積 20 日供貨量有較優的分類效力。此外兩實驗整體準確率為「10 日漲跌分類」較佳，推論較長時間的價格波動較容

易分類，但由於當日供貨量權重接逾兩實驗第一名，因此最終本研究只挑選當日供貨量作為供貨量唯一特徵。

5.3 天氣特徵

天氣特徵資料共有 19 個縣市，由於甘藍-初秋 LA1 生長週期約 50 至 60 天，因此本小節針對各個地區天氣特徵分成：「當日天氣」、「累積 30 日」、「累積 60 日」再獨立進行平均價漲跌三分類實驗。

此外由於天氣特徵於各縣市之天氣數值會因為地區相近使得不同縣市但有相似天氣特徵，又因為相似的特徵同時放入預測模型中不會使模型有更好的表現，且特徵越多所花費的時間和處理器運算成本越高，因此本節於漲跌分類實驗後、挑選特徵前加入各地天氣特徵關聯度視覺化分析，以挑選最少且多樣的天氣特徵。

5.3.1 溫度特徵

首先進行當日溫度之漲跌分類實驗，結果見表 5.3。

表 5.3 當日溫度實驗結果

| 當日溫度 | | 當日溫度 | |
|-------------------|-------|--------------------|-------|
| 3 日漲跌分類準確率 : 0.36 | | 10 日漲跌分類準確率 : 0.40 | |
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| 屏東縣 Temperature | 0.061 | 屏東縣 Temperature | 0.062 |
| 花蓮縣 Temperature | 0.060 | 南投縣 Temperature | 0.061 |
| 臺東縣 Temperature | 0.057 | 高雄市 Temperature | 0.059 |

| | | | |
|-----------------|-------|-----------------|-------|
| 南投縣 Temperature | 0.057 | 花蓮縣 Temperature | 0.059 |
| 高雄市 Temperature | 0.056 | 臺中市 Temperature | 0.058 |
| 臺中市 Temperature | 0.055 | 宜蘭縣 Temperature | 0.057 |
| 雲林縣 Temperature | 0.054 | 嘉義縣 Temperature | 0.055 |
| 臺北市 Temperature | 0.053 | 臺東縣 Temperature | 0.055 |
| 嘉義縣 Temperature | 0.052 | 臺南市 Temperature | 0.054 |
| 宜蘭縣 Temperature | 0.052 | 雲林縣 Temperature | 0.052 |
| 臺南市 Temperature | 0.051 | 苗栗縣 Temperature | 0.051 |
| 彰化縣 Temperature | 0.051 | 彰化縣 Temperature | 0.049 |
| 新竹縣 Temperature | 0.050 | 桃園市 Temperature | 0.049 |
| 桃園市 Temperature | 0.050 | 新竹縣 Temperature | 0.048 |
| 苗栗縣 Temperature | 0.049 | 臺北市 Temperature | 0.047 |
| 新北市 Temperature | 0.047 | 基隆市 Temperature | 0.045 |
| 嘉義市 Temperature | 0.046 | 新竹市 Temperature | 0.044 |
| 基隆市 Temperature | 0.046 | 嘉義市 Temperature | 0.044 |
| 新竹市 Temperature | 0.045 | 新北市 Temperature | 0.043 |

雖然見到糟糕準確率但符合預期結果：由於甘藍-初秋生長週期較長，因此各地當日溫度對於當日價格較無太大意義，故持續實驗累積溫度特徵，累積 30 日溫度實驗結果見表 5.4。

表 5.4 累積 30 日溫度實驗結果

| 累積 30 日溫度 3 日漲跌分類準確率：0.45 | | 累積 30 日溫度 10 日漲跌分類準確率：0.57 | |
|------------------------------|--------------|-------------------------------|--------------|
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| 南投縣 Temperature_30CNT | 0.058 | 南投縣 Temperature_30CNT | 0.061 |
| 嘉義縣 Temperature_30CNT | 0.057 | 高雄市 Temperature_30CNT | 0.059 |
| 屏東縣 Temperature_30CNT | 0.057 | 臺南市 Temperature_30CNT | 0.057 |
| 雲林縣 Temperature_30CNT | 0.056 | 基隆市 Temperature_30CNT | 0.057 |
| 高雄市 Temperature_30CNT | 0.056 | 臺東縣 Temperature_30CNT | 0.057 |
| 臺東縣 Temperature_30CNT | 0.054 | 宜蘭縣 Temperature_30CNT | 0.056 |
| 宜蘭縣 Temperature_30CNT | 0.053 | 花蓮縣 Temperature_30CNT | 0.055 |
| 嘉義市 Temperature_30CNT | 0.053 | 嘉義縣 Temperature_30CNT | 0.053 |
| 臺南市 Temperature_30CNT | 0.052 | 臺中市 Temperature_30CNT | 0.052 |
| 花蓮縣 Temperature_30CNT | 0.052 | 新竹市 Temperature_30CNT | 0.052 |
| 基隆市 Temperature_30CNT | 0.052 | 臺北市 Temperature_30CNT | 0.052 |
| 臺中市 Temperature_30CNT | 0.051 | 嘉義市 Temperature_30CNT | 0.051 |

| | | | |
|-----------------------|-------|-----------------------|--------------|
| 新竹市 Temperature_30CNT | 0.051 | 屏東縣 Temperature_30CNT | 0.048 |
| 桃園市 Temperature_30CNT | 0.050 | 新竹縣 Temperature_30CNT | 0.048 |
| 新竹縣 Temperature_30CNT | 0.050 | 苗栗縣 Temperature_30CNT | 0.047 |
| 臺北市 Temperature_30CNT | 0.048 | 雲林縣 Temperature_30CNT | 0.047 |
| 苗栗縣 Temperature_30CNT | 0.048 | 桃園市 Temperature_30CNT | 0.047 |
| 新北市 Temperature_30CNT | 0.047 | 新北市 Temperature_30CNT | 0.046 |
| 彰化縣 Temperature_30CNT | 0.045 | 彰化縣 Temperature_30CNT | 0.045 |

再計算候選縣市溫度特徵 Pearson 相關係數，並繪製圖 5.3。

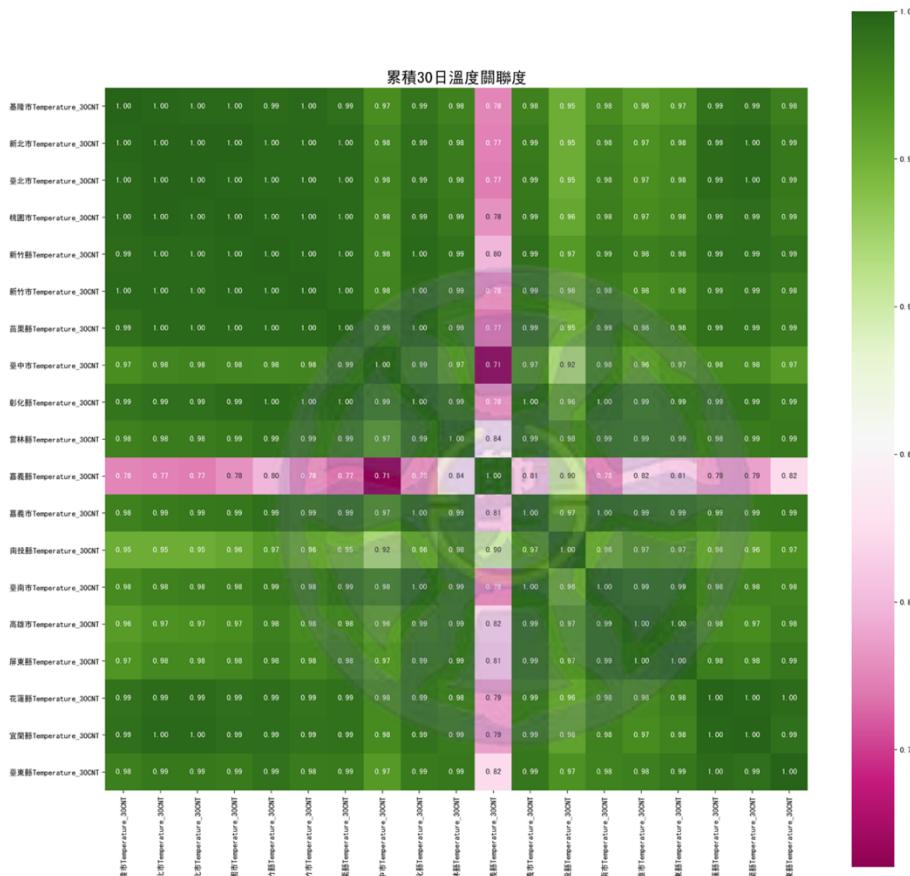


圖 5.3 累積 30 日溫度關聯度分析

「累積 30 日溫度實驗」結果準確率明顯高於「單日溫度實驗」，且嘉義、南投有較高的分類權重，但意外的是苗栗縣雖為主要產地之一，卻於「3 日漲跌分類」敬陪末座，因此進一步觀察圖 5.3 累積 30 日溫度關聯度分析，發現多數縣市溫度波動相近，但嘉義線相關度最低，其次為：台中、南投、高雄和屏

東，對於天氣特徵，本節將綜合：「甘藍之主要產地」、「價格漲跌分類權重」和「天氣特徵關聯度分析」挑選特徵，篩選策略為：甘藍-初秋 LA1 主要產地且分類權重高，但挑選的縣市彼此間關聯度低的天氣特徵。

累積 60 日溫度實驗結果見表 5.5。

表 5.5 累積 60 日溫度實驗結果

| 累積 60 日溫度 | | 累積 60 日溫度 | |
|-----------------------|--------------|-----------------------|--------------|
| 3 日漲跌分類準確率：0.45 | | 10 日漲跌分類準確率：0.60 | |
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| 嘉義縣 Temperature_60CNT | 0.061 | 嘉義縣 Temperature_60CNT | 0.067 |
| 南投縣 Temperature_60CNT | 0.058 | 南投縣 Temperature_60CNT | 0.064 |
| 臺東縣 Temperature_60CNT | 0.055 | 臺中市 Temperature_60CNT | 0.059 |
| 臺中市 Temperature_60CNT | 0.055 | 臺南市 Temperature_60CNT | 0.057 |
| 嘉義市 Temperature_60CNT | 0.054 | 新竹市 Temperature_60CNT | 0.055 |
| 臺南市 Temperature_60CNT | 0.054 | 臺東縣 Temperature_60CNT | 0.054 |
| 雲林縣 Temperature_60CNT | 0.053 | 高雄市 Temperature_60CNT | 0.053 |
| 桃園市 Temperature_60CNT | 0.053 | 嘉義市 Temperature_60CNT | 0.052 |
| 花蓮縣 Temperature_60CNT | 0.053 | 基隆市 Temperature_60CNT | 0.051 |
| 苗栗縣 Temperature_60CNT | 0.052 | 桃園市 Temperature_60CNT | 0.051 |
| 新竹市 Temperature_60CNT | 0.051 | 苗栗縣 Temperature_60CNT | 0.050 |
| 基隆市 Temperature_60CNT | 0.050 | 屏東縣 Temperature_60CNT | 0.050 |
| 屏東縣 Temperature_60CNT | 0.050 | 雲林縣 Temperature_60CNT | 0.050 |
| 新竹縣 Temperature_60CNT | 0.050 | 新竹縣 Temperature_60CNT | 0.049 |
| 高雄市 Temperature_60CNT | 0.050 | 花蓮縣 Temperature_60CNT | 0.047 |
| 宜蘭縣 Temperature_60CNT | 0.049 | 宜蘭縣 Temperature_60CNT | 0.047 |
| 臺北市 Temperature_60CNT | 0.048 | 新北市 Temperature_60CNT | 0.045 |
| 彰化縣 Temperature_60CNT | 0.048 | 彰化縣 Temperature_60CNT | 0.045 |
| 新北市 Temperature_60CNT | 0.048 | 臺北市 Temperature_60CNT | 0.044 |

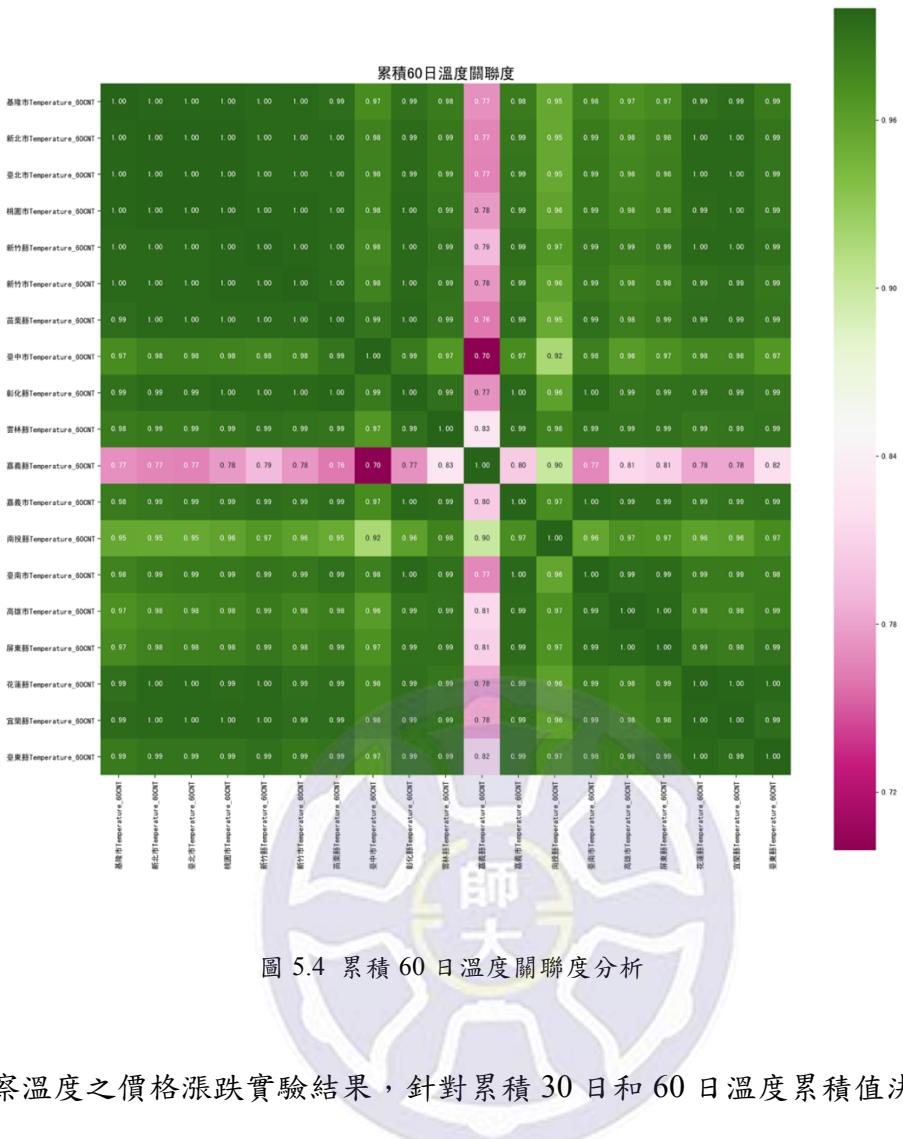


圖 5.4 累積 60 日溫度關聯度分析

觀察溫度之價格漲跌實驗結果，針對累積 30 日和 60 日溫度累積值決定個別選出四個縣市：嘉義縣、高雄市、雲林縣、南投縣，除此之外加上臺北市當日溫度，原因是考量台北當日溫度會影響市場交易。

5.3.2 降水量特徵

當日降水量實驗結果見表 5.6。

表 5.6 當日降水量實驗結果

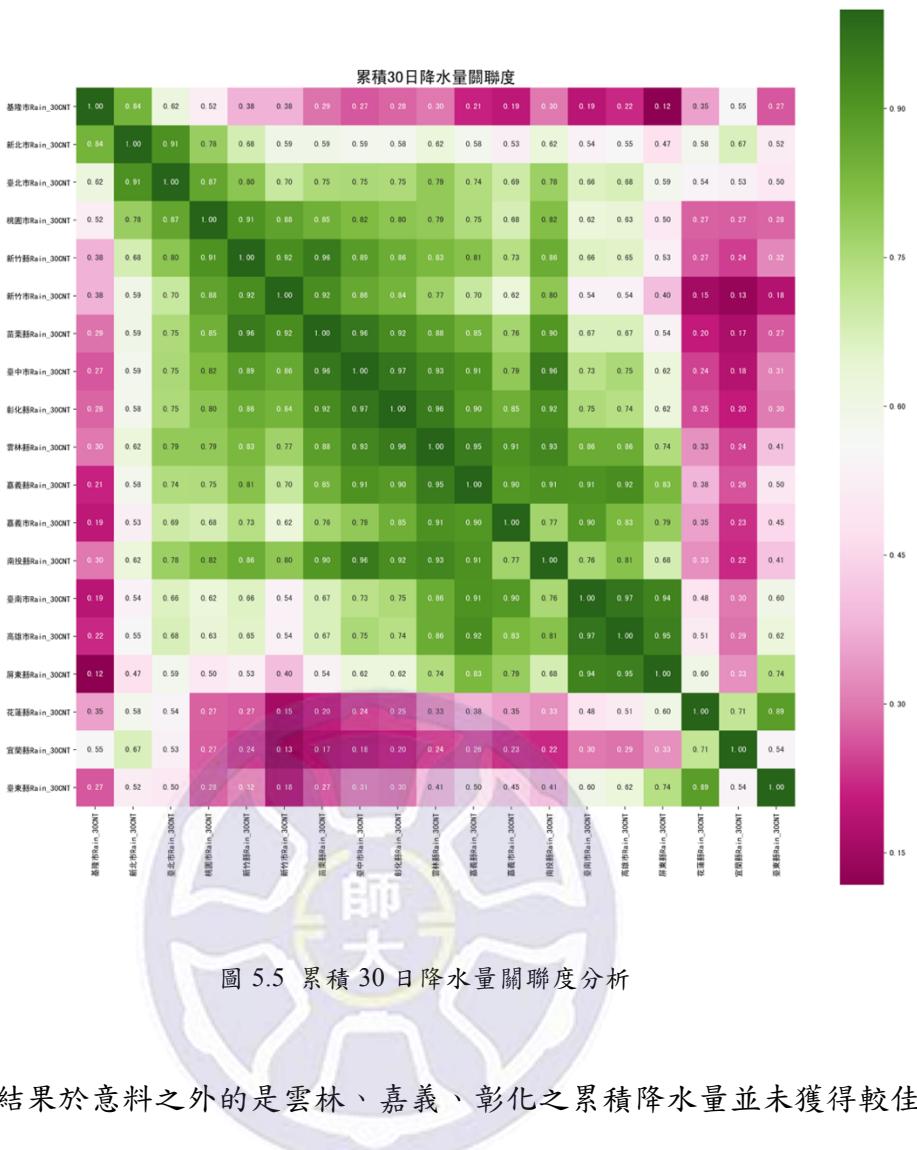
| 當日降水量 3 日漲跌分類準確率 : 0.37 | | 當日降水量 10 日漲跌分類準確率 : 0.36 | |
|----------------------------|-------|-----------------------------|-------|
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| 臺東縣 Rain | 0.083 | 花蓮縣 Rain | 0.083 |
| 花蓮縣 Rain | 0.083 | 臺東縣 Rain | 0.081 |
| 屏東縣 Rain | 0.076 | 宜蘭縣 Rain | 0.079 |
| 宜蘭縣 Rain | 0.075 | 屏東縣 Rain | 0.076 |
| 新北市 Rain | 0.064 | 南投縣 Rain | 0.066 |
| 高雄市 Rain | 0.063 | 新北市 Rain | 0.064 |
| 南投縣 Rain | 0.062 | 高雄市 Rain | 0.059 |
| 新竹縣 Rain | 0.054 | 臺北市 Rain | 0.057 |
| 臺北市 Rain | 0.054 | 苗栗縣 Rain | 0.052 |
| 苗栗縣 Rain | 0.050 | 新竹縣 Rain | 0.052 |
| 嘉義縣 Rain | 0.050 | 嘉義縣 Rain | 0.048 |
| 臺中市 Rain | 0.049 | 臺中市 Rain | 0.047 |
| 基隆市 Rain | 0.042 | 臺南市 Rain | 0.044 |
| 臺南市 Rain | 0.042 | 基隆市 Rain | 0.043 |
| 桃園市 Rain | 0.038 | 桃園市 Rain | 0.041 |
| 雲林縣 Rain | 0.033 | 雲林縣 Rain | 0.031 |
| 彰化縣 Rain | 0.028 | 彰化縣 Rain | 0.028 |
| 嘉義市 Rain | 0.025 | 嘉義市 Rain | 0.024 |
| 新竹市 Rain | 0.019 | 新竹市 Rain | 0.018 |

當日降水量實驗結果同當日溫度實驗，再繼續往下實驗，累積 30 日降水量

實驗結果見表 5.7。

表 5.7 累積 30 日降水量實驗結果

| 累積 30 日降水量 | | 累積 30 日降水量 | |
|-----------------------|--------------|-----------------------|--------------|
| 3 日漲跌分類準確率 : 0.46 | | 10 日漲跌分類準確率 : 0.59 | |
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| 宜蘭縣 Rain_30CNT | 0.061 | 宜蘭縣 Rain_30CNT | 0.062 |
| 臺東縣 Rain_30CNT | 0.060 | 花蓮縣 Rain_30CNT | 0.061 |
| 新北市 Rain_30CNT | 0.060 | 高雄市 Rain_30CNT | 0.058 |
| 花蓮縣 Rain_30CNT | 0.060 | 新北市 Rain_30CNT | 0.057 |
| 臺北市 Rain_30CNT | 0.058 | 臺東縣 Rain_30CNT | 0.056 |
| 基隆市 Rain_30CNT | 0.058 | 南投縣 Rain_30CNT | 0.054 |
| 高雄市 Rain_30CNT | 0.058 | 屏東縣 Rain_30CNT | 0.053 |
| 屏東縣 Rain_30CNT | 0.056 | 新竹縣 Rain_30CNT | 0.053 |
| 新竹縣 Rain_30CNT | 0.054 | 臺北市 Rain_30CNT | 0.053 |
| 南投縣 Rain_30CNT | 0.051 | 臺中市 Rain_30CNT | 0.053 |
| 桃園市 Rain_30CNT | 0.050 | 基隆市 Rain_30CNT | 0.052 |
| 臺中市 Rain_30CNT | 0.049 | 臺南市 Rain_30CNT | 0.052 |
| 嘉義縣 Rain_30CNT | 0.049 | 嘉義市 Rain_30CNT | 0.050 |
| 臺南市 Rain_30CNT | 0.049 | 苗栗縣 Rain_30CNT | 0.050 |
| 苗栗縣 Rain_30CNT | 0.047 | 嘉義縣 Rain_30CNT | 0.049 |
| 彰化縣 Rain_30CNT | 0.045 | 桃園市 Rain_30CNT | 0.049 |
| 嘉義市 Rain_30CNT | 0.043 | 雲林縣 Rain_30CNT | 0.044 |
| 雲林縣 Rain_30CNT | 0.042 | 新竹市 Rain_30CNT | 0.043 |
| 新竹市 Rain_30CNT | 0.038 | 彰化縣 Rain_30CNT | 0.042 |



實驗結果於意料之外的是雲林、嘉義、彰化之累積降水量並未獲得較佳分

類權重，但是其他甘藍主要產地宜蘭、南投仍有較佳表現，分析過後推論甘藍

現今主要為兩品種：甘藍-初秋 LA1 和甘藍改良種 LA2，前者主要為秋冬季

節，盛產地位於雲、嘉、彰；後者則在夏天產地位於宜、中、投，又因為夏天

整體產量下降，造成價格波動劇烈，故夏季產地的宜蘭、台中、南投之降水量

使得價格分類實驗表現較秋冬季節產地好。

繼續往下實驗，累積 60 日降水量實驗結果見表 5.8。

表 5.8 累積 60 日降水量實驗結果

| 累積 60 日降水量 3 日漲跌分類準確率 : 0.45 | | 累積 60 日降水量 10 日漲跌分類準確率 : 0.59 | |
|---------------------------------|--------------|----------------------------------|--------------|
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| 宜蘭縣 Rain_60CNT | 0.061 | 宜蘭縣 Rain_60CNT | 0.063 |
| 花蓮縣 Rain_60CNT | 0.061 | 南投縣 Rain_60CNT | 0.058 |
| 臺東縣 Rain_60CNT | 0.059 | 屏東縣 Rain_60CNT | 0.057 |
| 臺北市 Rain_60CNT | 0.058 | 新北市 Rain_60CNT | 0.056 |
| 新竹縣 Rain_60CNT | 0.057 | 臺東縣 Rain_60CNT | 0.056 |
| 屏東縣 Rain_60CNT | 0.056 | 苗栗縣 Rain_60CNT | 0.056 |
| 新北市 Rain_60CNT | 0.056 | 臺北市 Rain_60CNT | 0.056 |
| 基隆市 Rain_60CNT | 0.053 | 基隆市 Rain_60CNT | 0.053 |
| 南投縣 Rain_60CNT | 0.052 | 花蓮縣 Rain_60CNT | 0.053 |
| 桃園市 Rain_60CNT | 0.051 | 新竹縣 Rain_60CNT | 0.053 |
| 苗栗縣 Rain_60CNT | 0.051 | 臺中市 Rain_60CNT | 0.051 |
| 臺南市 Rain_60CNT | 0.051 | 桃園市 Rain_60CNT | 0.050 |
| 高雄市 Rain_60CNT | 0.051 | 高雄市 Rain_60CNT | 0.050 |
| 嘉義縣 Rain_60CNT | 0.050 | 新竹市 Rain_60CNT | 0.049 |
| 臺中市 Rain_60CNT | 0.048 | 嘉義縣 Rain_60CNT | 0.048 |
| 嘉義市 Rain_60CNT | 0.047 | 臺南市 Rain_60CNT | 0.047 |
| 雲林縣 Rain_60CNT | 0.045 | 彰化縣 Rain_60CNT | 0.046 |
| 彰化縣 Rain_60CNT | 0.042 | 雲林縣 Rain_60CNT | 0.045 |
| 新竹市 Rain_60CNT | 0.041 | 嘉義市 Rain_60CNT | 0.044 |

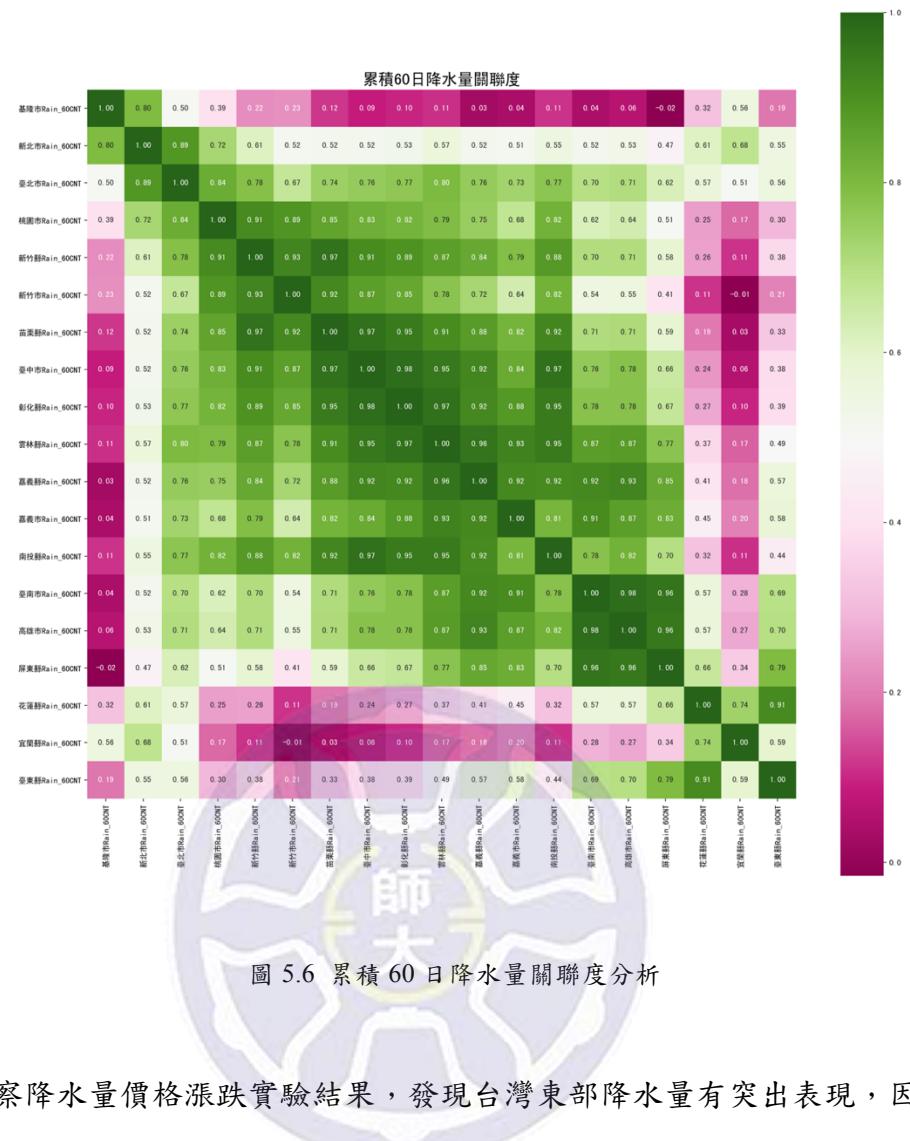


圖 5.6 累積 60 日降水量關聯度分析

觀察降水量價格漲跌實驗結果，發現台灣東部降水量有突出表現，因此決定加入東部一縣市：花蓮縣，針對兩實驗個別挑選四個縣市：

- 累積 30 日降水量：宜蘭縣、花蓮縣、南投縣、雲林縣。
- 累積 60 日降水量：宜蘭縣、花蓮縣、南投縣、苗栗縣。
- 除此之外加上臺北市當日降水量，原因同溫度特徵。

綜合 2 項天氣特徵對於甘藍-初秋 LA1 平均價漲跌分類實驗，經過多次分類

測試發現所有實驗皆於「10 日價格漲跌分類」有較佳準確率約 50% 至 60%，此外於降水量實驗中寶小節額外進行「平均 30 日降水量實驗」、「平均 60 日降

水量實驗」但結果與累積實驗相似，因此決定溫度與降水量皆採累積方式之特徵變換。

5.4 颱風警報發布特徵

颱風是影響菜價很大的原因，過去甚至發生颱風生成到警報發布後菜價隨即攀升，但是最後颱風並未登陸，甚至無實質影響，可見人民與供貨商之預期心理造成菜價震盪，因為營業店家或是大眾擔心颱風襲擊後菜價漲價甚至價格居高不下，於是選擇先前購買更多蔬菜備貨提防。

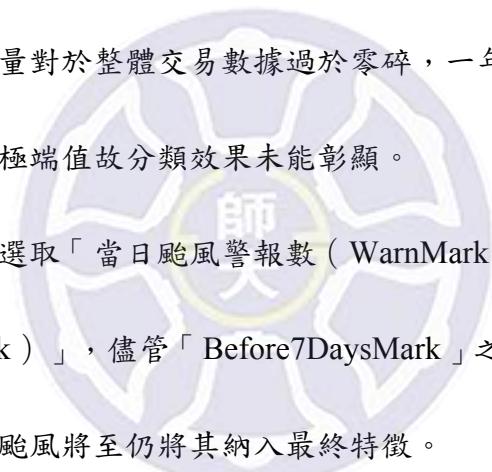
本研究擷取颱風警報資料僅能獲得警報當日的訊息，但事實上颱風在生成至發佈警報前，新聞皆會報導讓民眾提前了解及防範，因此為了模擬先前便知道颱風將至，因此平移颱風警報特徵，轉換成單位天數前知道颱風將至，此處單位天數定為 1 至 15 日，再進行颱風警報特徵實驗，結果見表 5.9。

表 5.9 颱風警報特徵實驗結果

| 颱風警報特徵 3 日漲跌分類準確率：0.38 | | 颱風警報特徵 10 日漲跌分類準確率：0.37 | |
|---------------------------|--------------|----------------------------|--------------|
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| WarnMark | 0.354 | WarnMark | 0.598 |
| Before13DaysMark | 0.066 | Before1DaysMark | 0.047 |
| Before10DaysMark | 0.066 | Before8DaysMark | 0.037 |
| Before11DaysMark | 0.065 | Before5DaysMark | 0.037 |
| Before3DaysMark | 0.063 | Before7DaysMark | 0.035 |
| Before1DaysMark | 0.050 | Before15DaysMark | 0.034 |
| Before9DaysMark | 0.048 | Before10DaysMark | 0.031 |
| Before8DaysMark | 0.045 | Before11DaysMark | 0.028 |

| | | | |
|------------------------|--------------|------------------|-------|
| Before6DaysMark | 0.043 | Before12DaysMark | 0.028 |
| Before5DaysMark | 0.037 | Before3DaysMark | 0.027 |
| Before4DaysMark | 0.037 | Before2DaysMark | 0.024 |
| Before7DaysMark | 0.033 | Before13DaysMark | 0.017 |
| Before12DaysMark | 0.030 | Before4DaysMark | 0.015 |
| Before2DaysMark | 0.026 | Before14DaysMark | 0.011 |
| Before14DaysMark | 0.018 | Before9DaysMark | 0.010 |
| Before15DaysMark | 0.011 | Before6DaysMark | 0.010 |

由實驗得知分類效果最好的仍為當日颱風警報數：「 WarnMark 」，其次為未來 5 至 13 天後之颱風警報數，推斷應為颱風警報前兩週，民眾於新聞報導得知颱風將至，因而產生預期心理。但是實驗結果準確率不到四成，其可能原因為颱風警報資料量對於整體交易數據過於零碎，一年內可能只有十來次發布颱風警報數，屬於極端值故分類效果未能彰顯。



颱風特徵最終選取「當日颱風警報數（WarnMark）」及「七日前颱風警報（Before7DaysMark）」，儘管「Before7DaysMark」之分類權重不高，但考量現實大眾會提早得知颱風將至仍將其納入最終特徵。

5.5 市場休息與日期特徵

市場休息特徵包含：前一天市場休息「YesterDayRest」、隔天市場休息：「TomorrowRest」、前一天連休幾日「YesterDayRestCNT」、隔天連休幾日「TomorrowRestCNT」。日期特徵包含星期「WeekDay」、月份「Month」，由於此兩類別特徵數量較少，因此合併後進行實驗，結果表 5.10。

表 5.10 休市與日期特徵實驗結果

| 休市與日期特徵 3 日漲跌分類準確率：0.40 | | 休市與日期特徵 10 日漲跌分類準確率：0.35 | |
|----------------------------|------|-----------------------------|------|
| 特徵 | 分類權重 | 特徵 | 分類權重 |
| Month | 0.54 | Month | 0.57 |
| WeekDay | 0.29 | WeekDay | 0.26 |
| TomorrowRestCNT | 0.06 | TomorrowRestCNT | 0.06 |
| YesterDayRestCNT | 0.05 | YesterDayRestCNT | 0.05 |
| TomorrowRest | 0.02 | TomorrowRest | 0.01 |
| YesterDayRest | 0.01 | YesterDayRest | 0.01 |

本研究加入日期特徵考量為台北一市場固定於星期一休市，且星期五為假

日前一天，一般民眾和店家會於星期五購買更多蔬菜，農產品需求較大，因此將日期特徵納入實驗模型，合併市場休息特徵與日期特徵可發現市場休息特徵分類權重較低，推論其原因可能類似於颱風警報即市場休息對於交易數據集過於零散，但考量現實生活仍將這些候選特徵全部選為最終特徵。

本章經由上述實驗觀察不同類型的候選特徵對於價格漲跌分類的權重度，並按照現實生活經驗和候選特徵彼此關聯度，逐步挑選出價格預測模型的最終特徵再按照日期合併，合併後包含日期共 31 特徵共 2352 天，扣除休市日實際資料量為 1959 筆。

第六章、價格預測模型

本研究之價格預測模型使用 LSTM 長短期記憶模型，本研究針對國產最大宗蔬菜：甘藍-初秋 LA1，藉由特徵工程清理資料和特徵分析最終篩選出 31 個特徵於本章進行模型訓練，資料量共 1959 天資料，模型訓練前將 2012 至 2017 年之特徵資料集作為訓練集（共 1812 筆）並以 2018 年作為測試集（共 147 筆）。

6.1 預測未來 5 日價格模型

本研究之價格預測模型初期實驗目標為：輸入連續 5 日特徵資料預測未來連續 5 日價格（5-D_LSTM）。模型架構為 3 層 LSTM、1 層輸出層。經過多次實驗與參數調整，最終網路架構參數為：神經元數量為 200、dropout 為 0.2，Learning Rate 為 0.005、模型優化演算法為 Mini-batch Gradient Descent (SGD)、誤差計算方法為 Mean squared error (MSE)，此外訓練前會先將數值縮小以利於神經網路訓練，方法為一般化 (Normalization)，使得每一筆資料平均值為 0、標準差為 1，最後將篩選出的 31 個特徵投入模型內並重複訓練 200 次。5-D_LSTM 模型訓練結果之 Mean squared error 折線圖見圖 6.1。

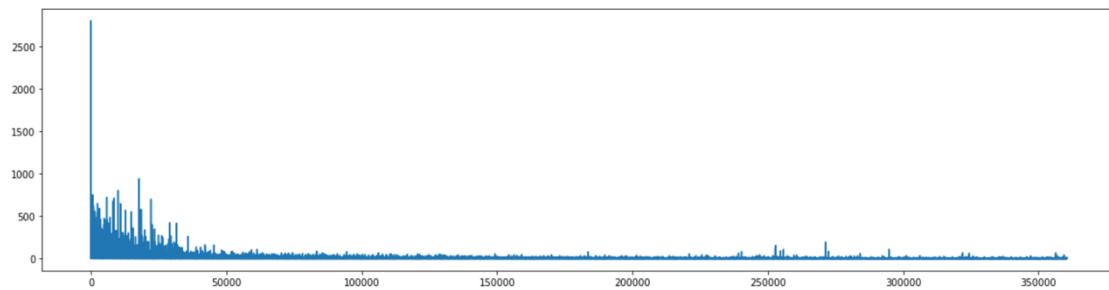


圖 6.1 5-D_LSTM -MSE 折線圖

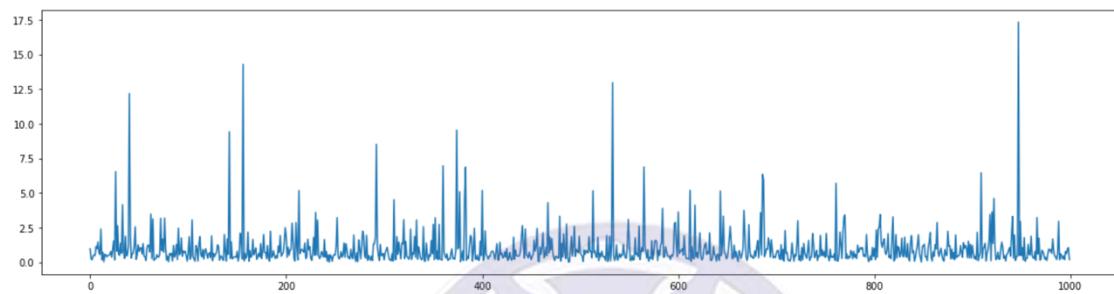


圖 6.2 5-D_LSTM -MSE 折線圖-訓練最後 1000 次

經過多次模型的調整，錯誤率的表現最終於 2 至 0 之間震盪，但觀察圖 6.2

錯誤率不時會漲至 10 至 17 之間。接著將測試集資料進行預測實驗，結果見圖

6.3。縱軸為價格，橫軸為時間，藍色為實際價格走向，橘色為 5-D_LSTM 價格預測模型結果，並將誤差結果繪成折線圖（圖 6.4）。

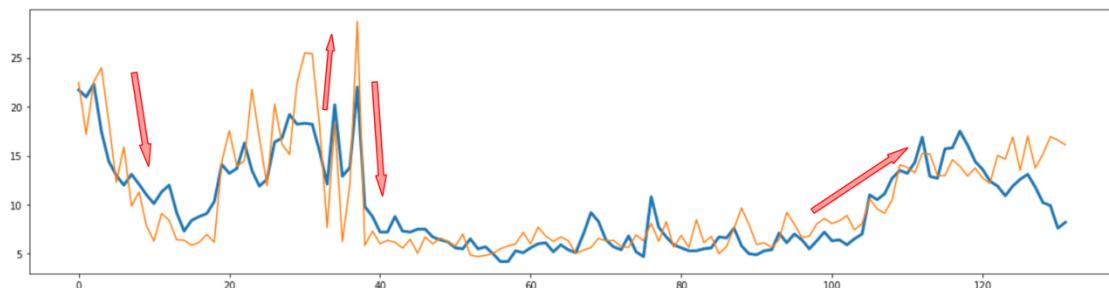


圖 6.3 5-D_LSTM 測試集-預測結果

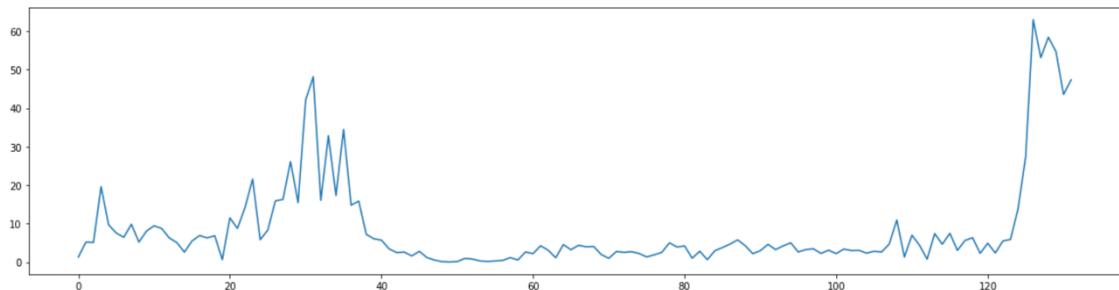


圖 6.4 5-D_LSTM 測試集-MSE 折線圖

觀察圖 6.3 結果發現此價格預測模型的結果於未來「整體趨勢」有較佳的預測結果，但是對於短期的價格漲跌之可信度仍不穩定；觀察圖 6.4 預測結果之 MSE 值有些會飆升至 30 以上，但是參照預測結果可以發現雖然預測價格脫離實際價格許多，但是整體的漲跌趨勢是正確的，意即雖然實際預測值不準確但是漲跌的趨勢有符合實際結果，不過還是仍有表現不佳的現象，於此再提出其他不同日期長度的價格預測模型，期望藉由觀察多個模型結果來提高推測預測價格的準確度。

6.2 價格預測混合模型

為了提高價格預測可信度，本研究加入不同日期範圍的價格預測模型，分別為預測未來連續 3 日（3-D_LSTM）、10 日（10-D_LSTM）、15 日（15-D_LSTM），包含上節共四個模型，新建的的類神經網路模型參數設定與前一節之 5-D_LSTM 相同，所有價格預測模型訓練完成後，將測試資料集分別進行四

模型價格預測以觀察模型之準確度並將結果會成以下圖形，粗藍線為實際價格，橘線為預測結果。

3-D_LSTM 預測結果見圖 6.5，可以觀察預測出大幅度的價格漲跌清晰準確，但是小幅度的價格震盪如紅圈處，在預測結果上並不理想，誤差結果（圖 6.6）同 5-D_LSTM：實際預測值不準確但是漲跌的趨勢有符合實際結果。

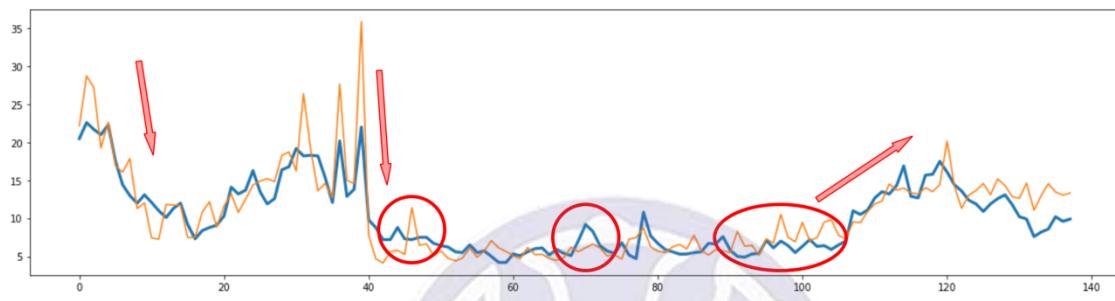


圖 6.5 3-D_LSTM 測試集-預測結果

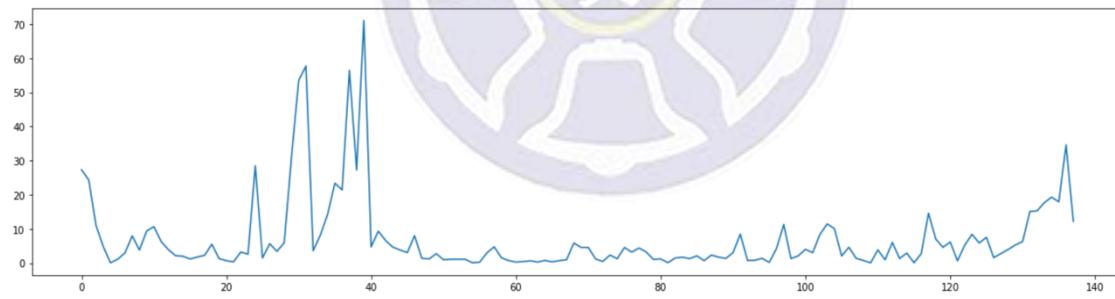


圖 6.6 3-D_LSTM 測試集-MSE 折線圖

10-D_LSTM 預測結果見圖 6.7，觀察可發現漲跌趨勢不如 3-D_LSTM、5-D_LSTM 精確，但是觀察誤差結果（圖 6.8）卻是表現最好，於四模型整體中得到最小誤差。

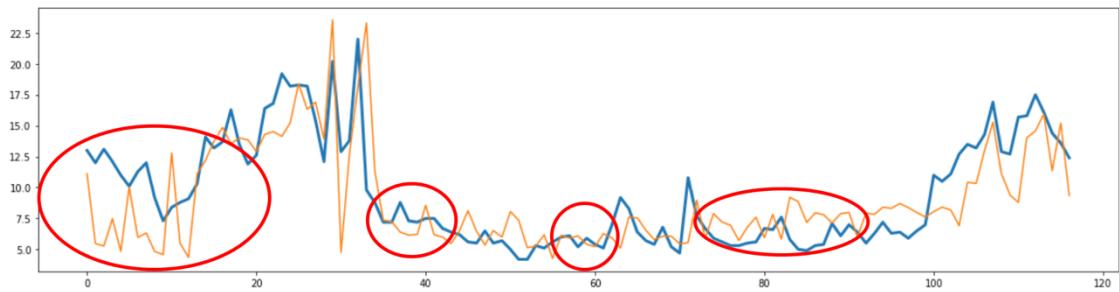


圖 6.7 10-D_LSTM 測試集-預測結果

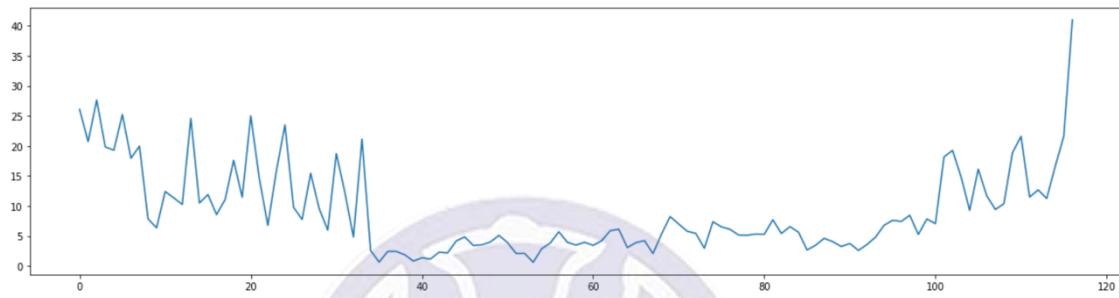


圖 6.8 10-D_LSTM 測試集-MSE 折線圖

15-D_LSTM 預測結果見圖 6.9，預測結果的波形與 10-D_LSTM 相較更無太大的起伏，由誤差結果（圖 6.10）發現預測結果相較於其他三模型偏離實際價格更多，不過假若單看預測價格的走勢卻與實際價格有相似的波形。

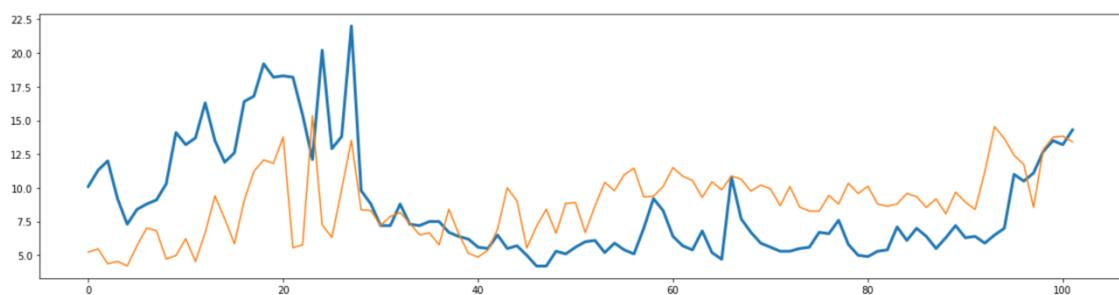


圖 6.9 15-D_LSTM 測試集-預測結果

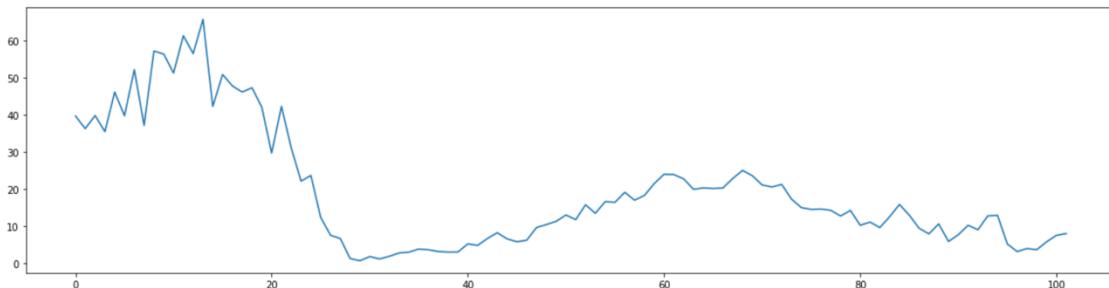


圖 6.10 15-D_LSTM 測試集- MSE 折線圖

觀察四個模型預測的結果可得以下不同模型的特性：

1. 四模型對於實際價格出現大幅度漲跌皆有良好預測表現。
2. 短時間價格預測模型（3、5-D_LSTM）結果相較於長時間模型預測結果敏銳且精確，但是對於價格平穩期卻有失準現象。
3. 長時間價格預測模型（10、15-D_LSTM）之預測結果偏離實際價格，但是 10-D_LSTM 於價格平穩期之預測波形比短時間價格預測模型更準確，且 15-D_LSTM 現階段來看預測效力最差，可能無法有好的貢獻。

綜合以上觀察與分析得知本研究之價格預測混合模型於實際應用需配合實際價格波動範圍，再按照不同模型特性交互觀察才能得出預測結論。

接著進行模擬預測實驗：隨機挑選日期進行價格預測，並將結果以圖表呈現：藍色粗線段為真實價格；橘色線段為 3-D_LSTM；綠色線段為 5-D_LSTM；紅色線段為 10-D_LSTM；紫色線段為 15-D_LSTM。

觀察隨機實驗 1（圖 6.11），前三日預測結果為 3、10-D_LSTM 兩模型較準確，且 3-D_LSTM 有準確的符合實際價格波動；三至五日結果中 10-D_LSTM

為錯誤跌價趨勢；五至十日結果中 10-D_LSTM 結果波形相對準確，但 15-D_LSTM 却有錯誤的漲跌趨勢；十日之後實際價格平緩，15-D_LSTM 之預測準確。整體來看本次隨機實驗之實際價格漲跌起於 10 元內，算是小幅度的價格漲跌，綜合來看 3、10-D_LSTM 兩模型有較佳表現。

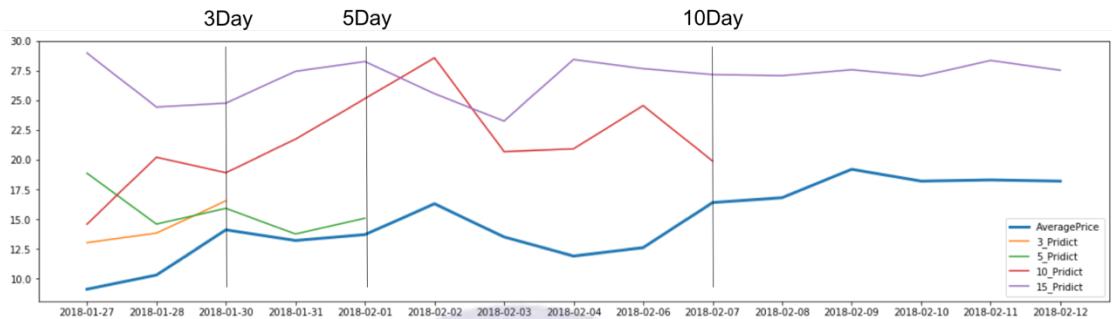


圖 6.11 價格預測混合模型-隨機實驗 1

觀察隨機實驗 2（圖 6.12），本次隨機實驗之實際價格比隨機實驗 1 的波動更小，本次隨機實驗之四模型幾乎沒有較好的預測結果，推論原因該時期價格穩定，加上本研究之價格預測隨機實驗於小幅度之價格震盪表現較差。

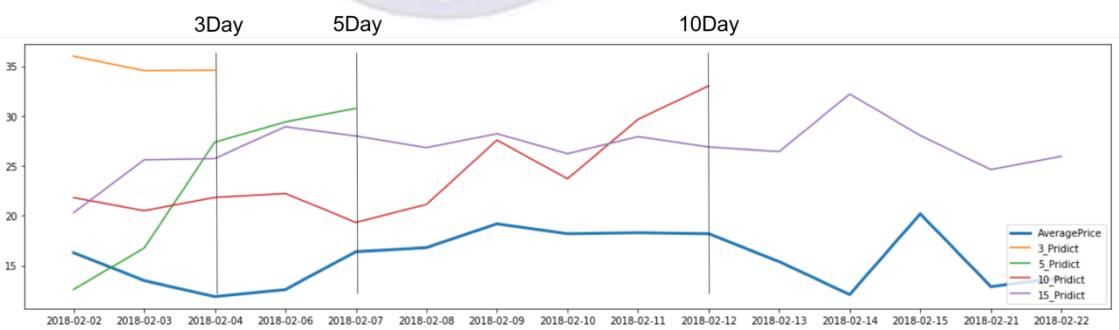


圖 6.12 價格預測混合模型-隨機實驗 2

觀察隨機實驗 3 (圖 6.13)，第二至三日內價格大漲約 10 元，除了 15-D_LSTM 之外其他三模型皆有正確漲價趨勢，如 3-D_LSTM 相當準確；5-D_LSTM 雖然結果偏離實際價格但預測之波動也符合實際價格走向，且第五至六日也有正確的跌價趨勢，五日後 10、15-D_LSTM 兩模型於價格趨於穩定間也並無顯著的漲跌誤判，綜合來看為較成功的價格預測實驗。

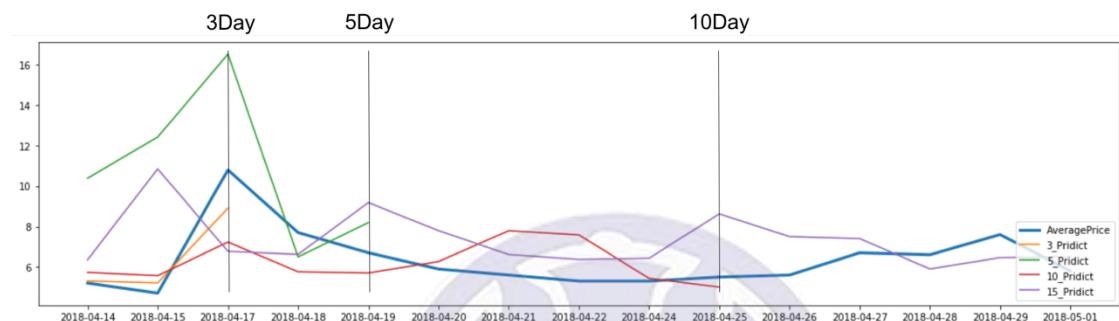


圖 6.13 價格預測混合模型-隨機實驗 3

本小節綜合以上實驗發現若價格漲跌起伏平穩的情境時預測效果較差，對應到現實生活通常為供貨量穩定如 12、1、2 月等。此外短時間價格預測模型 (3、5-D_LSTM) 有較好的預測表現，不過微觀來看各個模型皆有其優劣，因此本研究最終仍保留四個模型，期望於實際應用時能夠交互參照此混合預測模型之四個結果進行價格預測之決策。

6.3 價格預測模型結果討論

本研究從資料擷取搜集農產品價格相關資料，並藉由資料視覺化來討論價格變動的影響因素，再透過特徵分析篩選不同種類的候選特徵對於價格漲跌分

類實驗的權重度，並按照現實生活經驗和候選特徵彼此關聯度選出對於價格漲跌影響力最高的特徵，最終建立出價格預測混合模型包含：3、5、10、15-D_LSTM 四個子模型，分別用來預測未來 3、5、10、15 日的甘藍價格，本研究期望於實際應用上能夠藉由綜合觀察四個預測模型的結果並互參照得出未來價格漲跌情形。

價格預測模型訓練過程由圖 6.1 Mean squared error 折線圖可見訓練末期誤差值於 2 至 0 之間震盪，理想上神經網路訓練之 MSE 誤差值要接近 0，但由於價格變動的因素太多，礙於時間有限，本研究無法一一搜集所有影響價格因素來進行實驗，僅能以影響最大的天氣特徵及市場交易訊息進行分析，誤差值雖然無法持續穩定的下降，但相較過往其他參數已算表現最好。

實驗最終的混合模型預測對於實際應用來說還有很大的改進空間，對於價格預估的錯誤率本研究實驗至此有以下延伸討論：

1. 搜集更多特徵進行分析與實驗訓練如油價。或是針對現有特徵進行更深度的特徵工程，如將天氣特徵建構出連續狀態特徵，如是否連日下雨超過 20 毫升；連日溫度超過 30 度；合併溫度與降水量建構新特徵：當日溫度高於 30 度且降下大雨（由於高溫與積水容易使植物敗壞）。
2. 承 1，除了其他農產品市場之外，近年來各大零售或量販通路也包含農產品的銷售，因此對於農產品交易的研究可能也需採納各大通路之價

格與供貨量進行分析。

3. 本研究架構為價格金額回歸預測，對於現實生活議題預測，加上資料量不多的前提下其實很難有穩定優異的預測結果，但可以調整預測方式再實施混合模型架構以進行預測，例：針對價格之漲跌與否進行分類預測，預測判斷當日價格於 3 日內之漲跌是否超出 0.5、1、2 個離散程度，由回歸預測轉換為分類預測，再綜合混合模型已達到更精準價格預測。

4. 由於類神經網路需要大量的資料和多次的訓練，因此未來可以搜集更多資料並持續改善模型參數、提高訓練次數以提高預測準確度。

6.4 研究結果與未來展望

本研究使用開放資料作為特徵進行預測甘藍-初秋 LA1 批發價，前段大量的特徵工程需要資料清洗與偵錯，研究結果希望能獲得影響農產品批發價的主要影響特徵、預測未來批發價範圍區間。當系統獲得有價值的菜價漲幅條件後，為農產品之交易資訊透明化做出貢獻：

1. 農民：使農民交易時對於自己的商品價格更有信心。
2. 行口、中盤批發商：使批發商採買及販售時有價額外訊息幫助做決策，作為進貨量依據，降低不適當進貨量造成額外的運輸、搬運、整理成本。未來更能夠將本研究開發成自動化和即時網頁或行動裝置之

應用程式。

3. 開放資料：對於資料的搜集、整理及視覺化提出本研究的方法與流程，程式撰寫皆以操作、分析大量資料為方向，因此對於其他品項的蔬菜水果或花卉，皆能直接利用本研究之程式功能進行繪圖、分析與觀察。

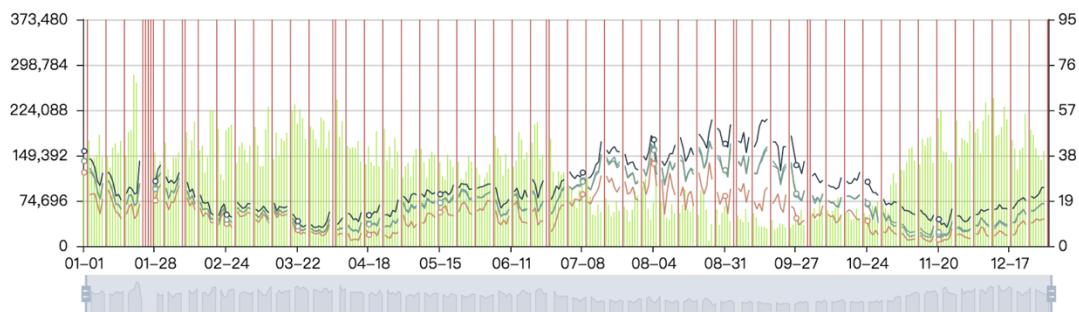
4. 特徵篩選方法：本研究提出農產品價格漲跌特徵之篩選提出兩大方
法：第四章使用多項資料視覺化和關聯度分析來觀察農產品各品項於
不同情境之交易訊息分析，呈現出大量資料內部之關係並針對各個現
象進行討論，再藉由第五章特徵分析進行價格預測特徵之實驗最終挑
選出價格影響力高之目標特徵。

儘管本研究目標僅針對台北一市場的甘藍進行研究，但全國蔬菜品項共 386
項（國產蔬菜 349 項），全部農產品品項共 2154 項，期望未來農產品價格議題
能有朝更多品項或是更深度的研究。

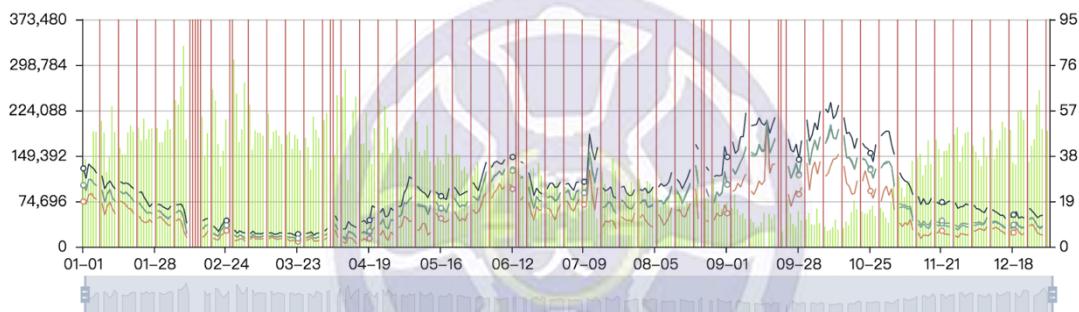
附錄

本研究目標蔬菜：甘藍-初秋 LA1 之交易訊息分析：

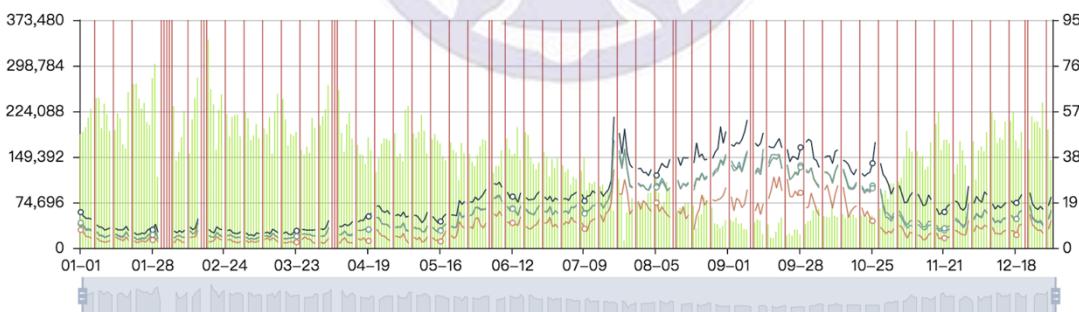
2012



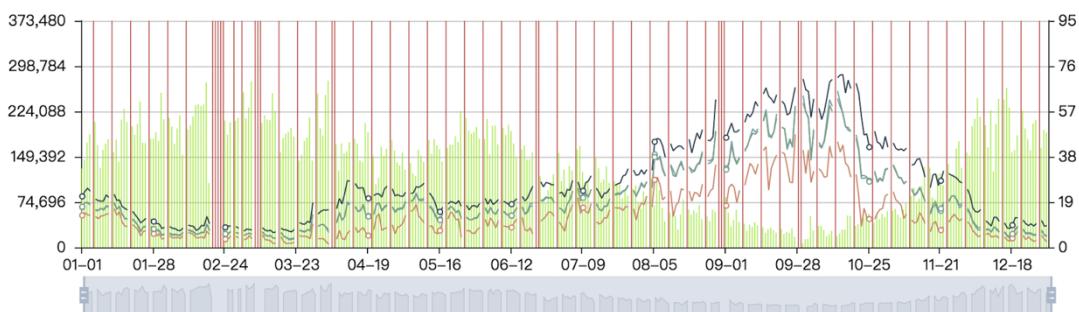
2013



2014

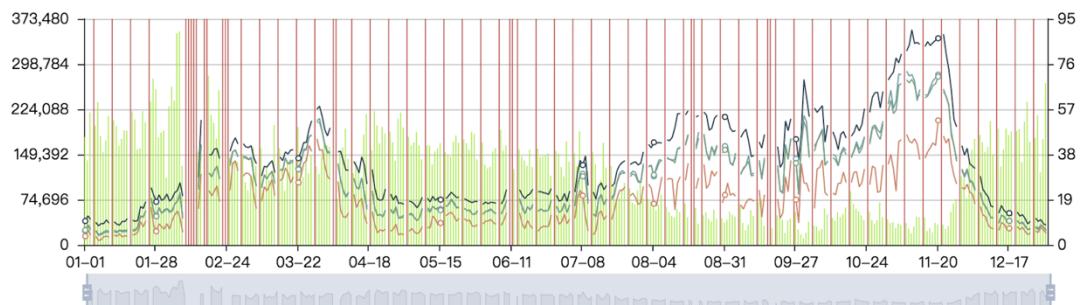


2015



2016

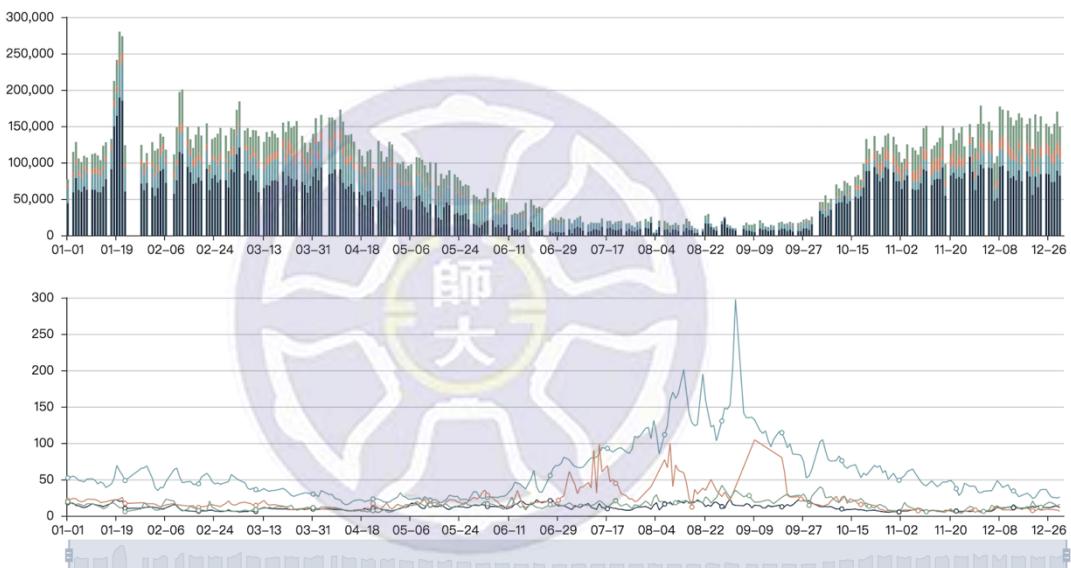
供貨量 (Green) 休市 (Red) BestPrice (Yellow) BetterPrice (Blue) GoodPrice (Orange) AveragePrice (Grey)



甘藍-初秋 LA1 供貨量高相關分析：

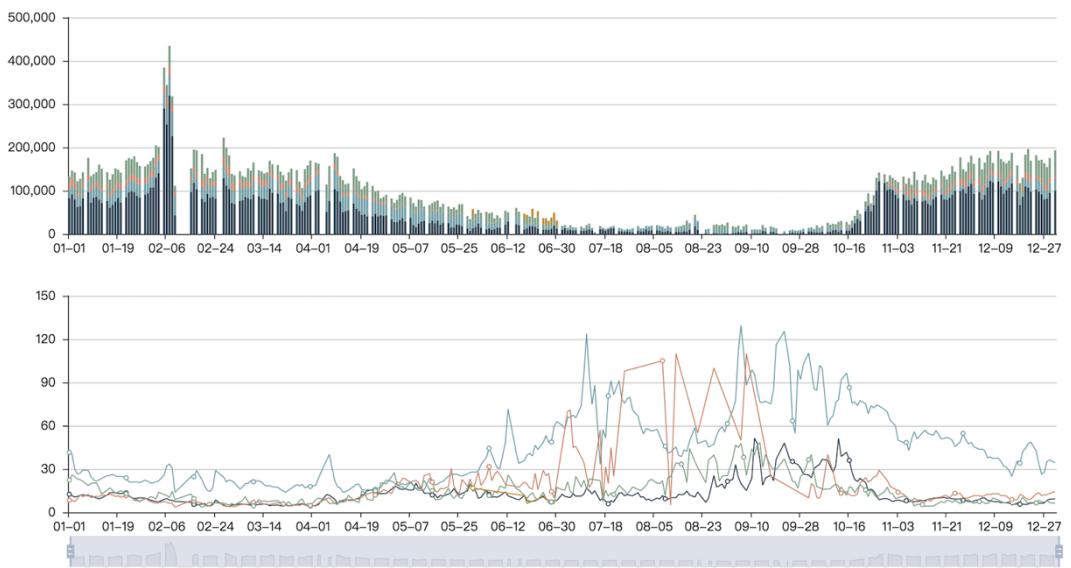
2012

甘藍-初秋 LA1 蘿蔔-研仔 SA32 青蔥-粉蔥 SE6 球莖甘藍 SW1 包心白菜-包頭蓮 LC3 蘿蔔-研仔 SA3



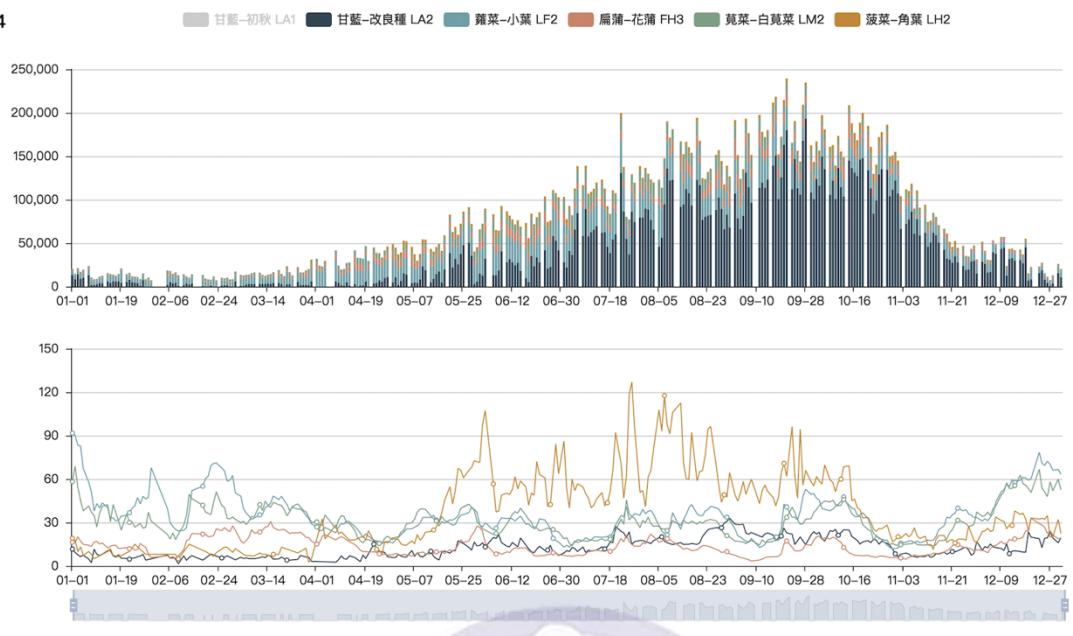
2013

甘藍-初秋 LA1 蘿蔔-研仔 SA32 青蔥-粉蔥 SE6 球莖甘藍 SW1 包心白菜-包頭蓮 LC3 蘿蔔-研仔 SA3

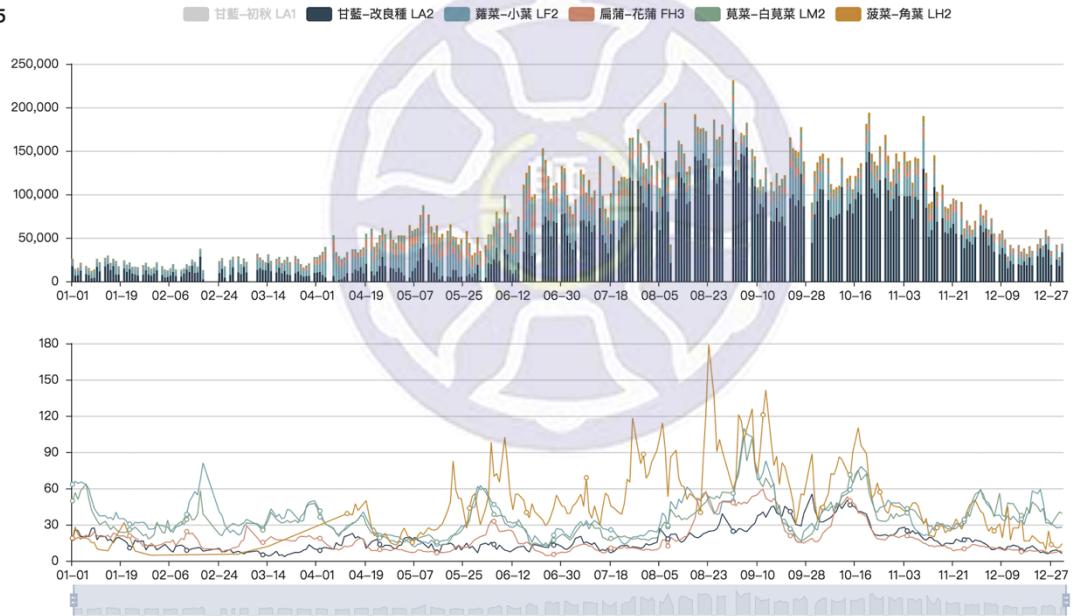


甘藍-初秋 LA1 供貨量低相關分析：

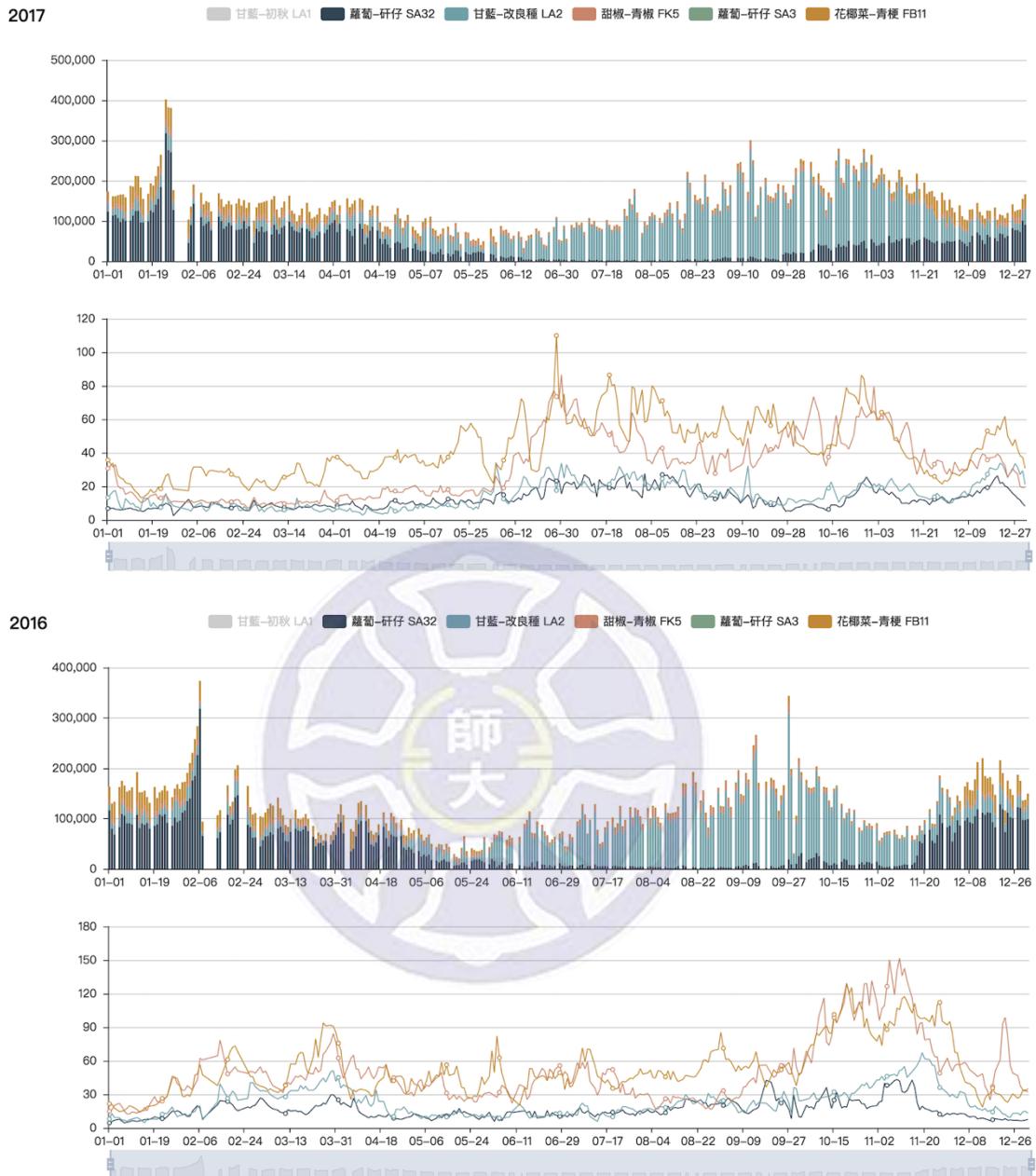
2014



2015

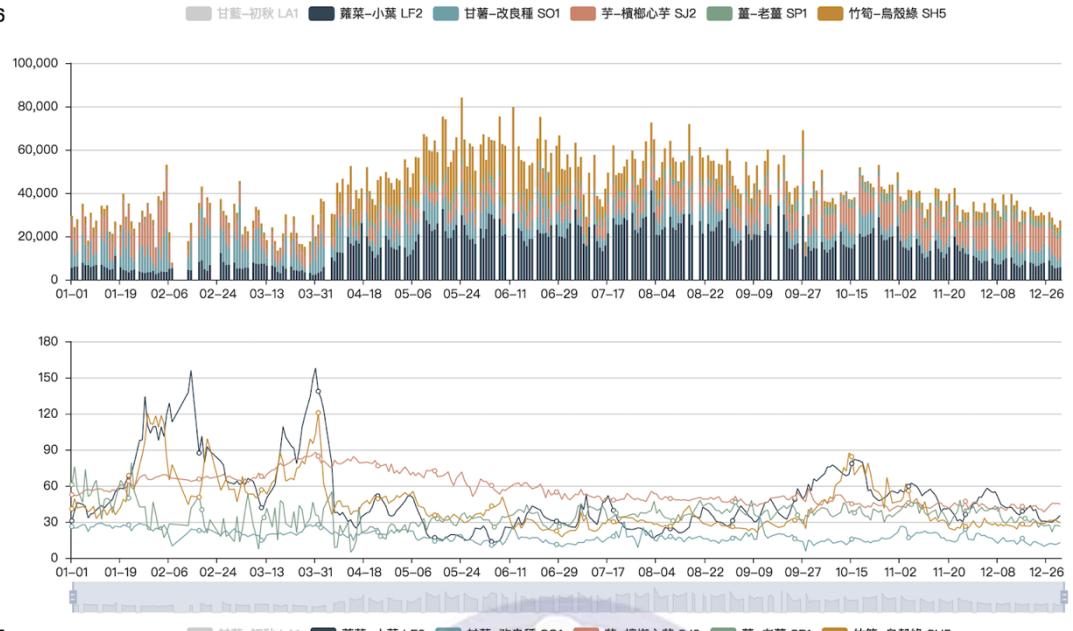


甘藍-初秋 LA1 平均價高相關分析：

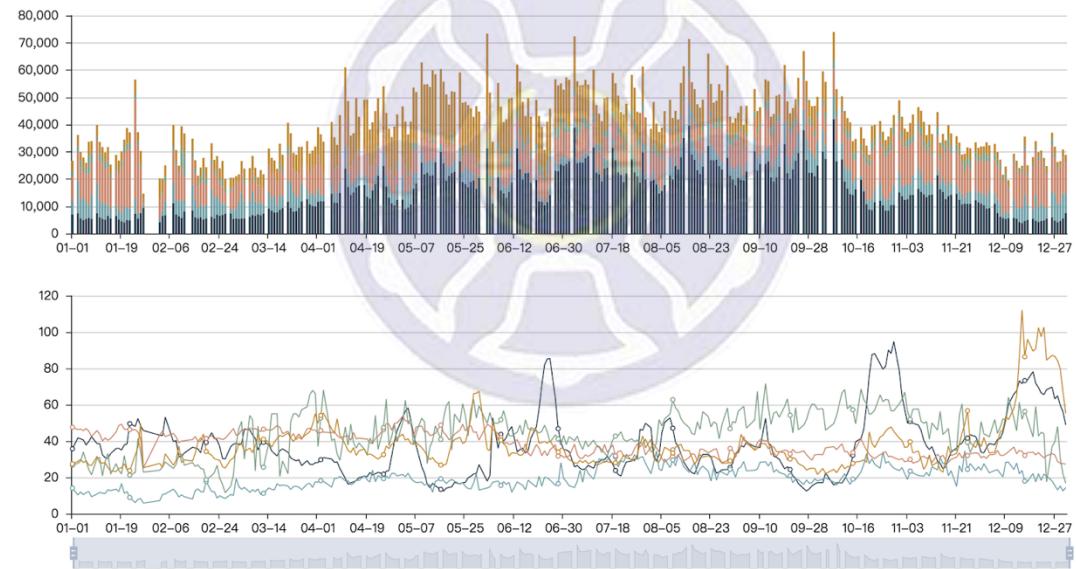


甘藍-初秋 LA1 平均價低相關分析：

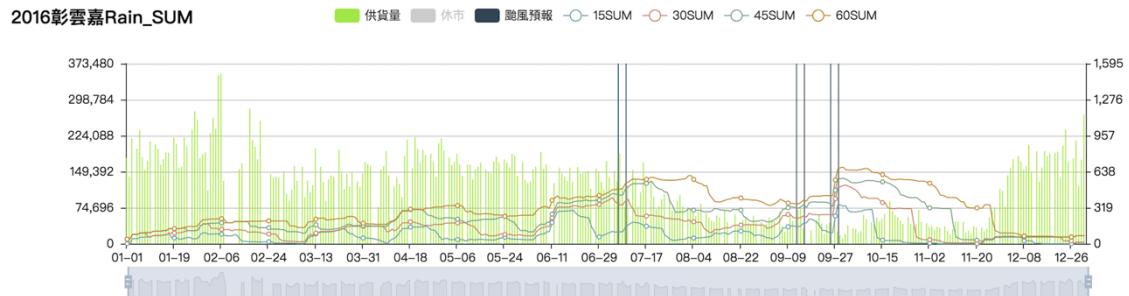
2016



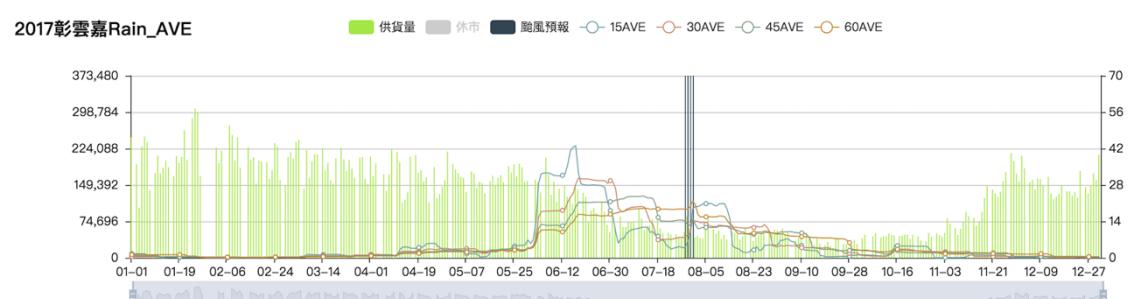
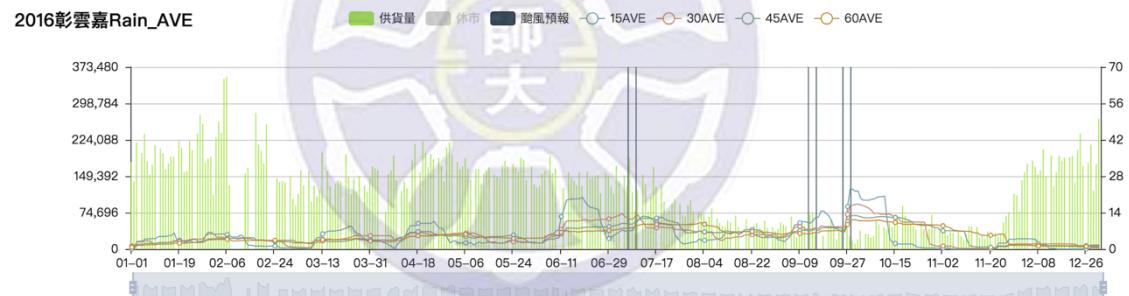
2017



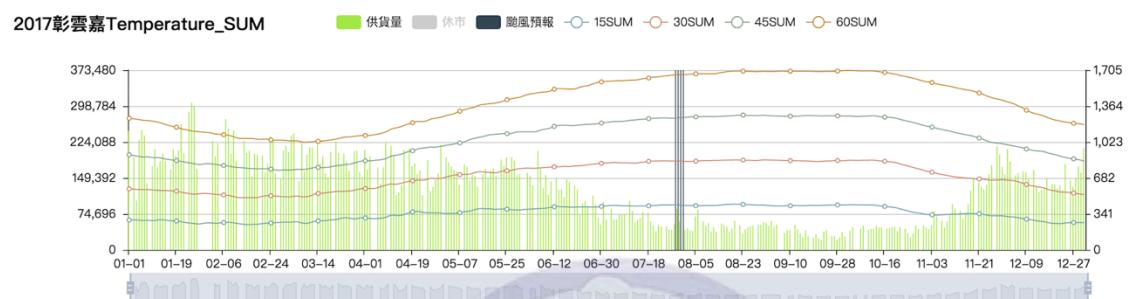
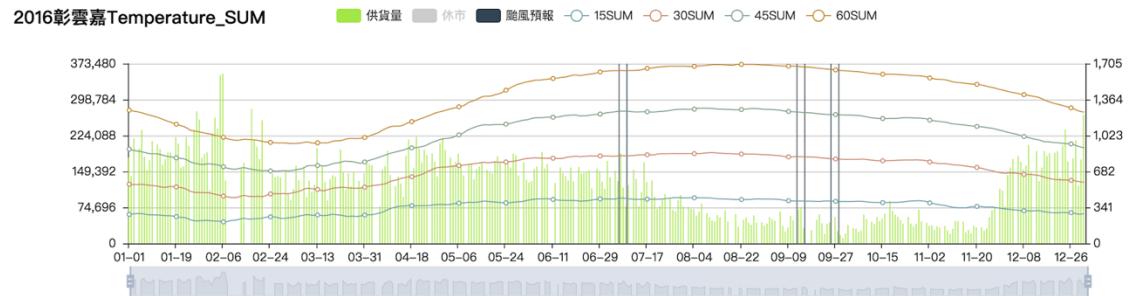
彰化、雲林、嘉義累積雨量分析：



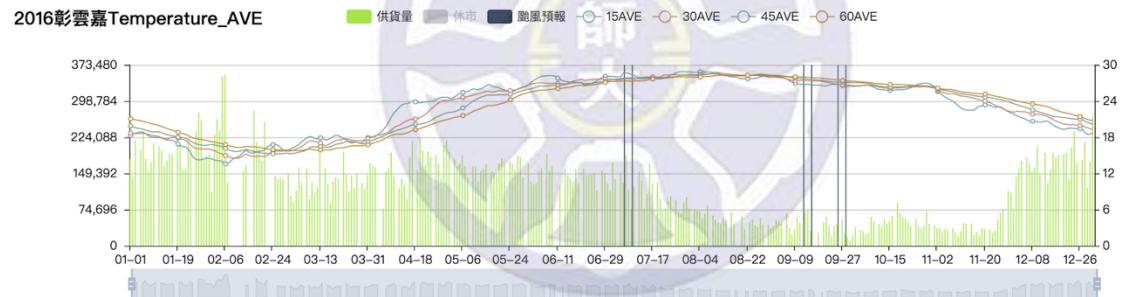
彰化、雲林、嘉義平均雨量分析：



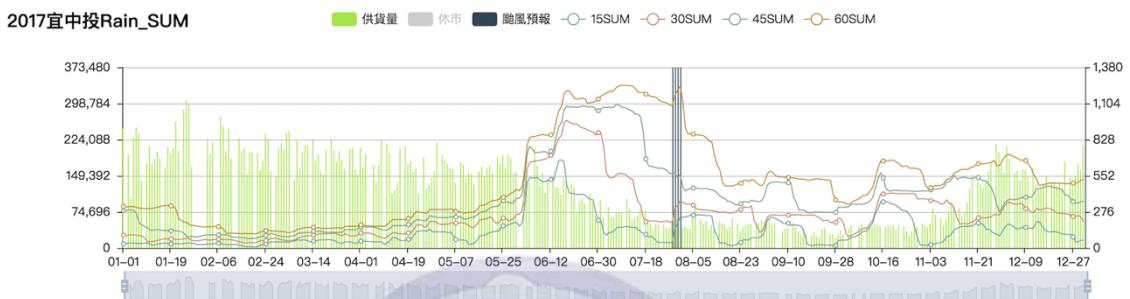
彰化、雲林、嘉義累積溫度分析：



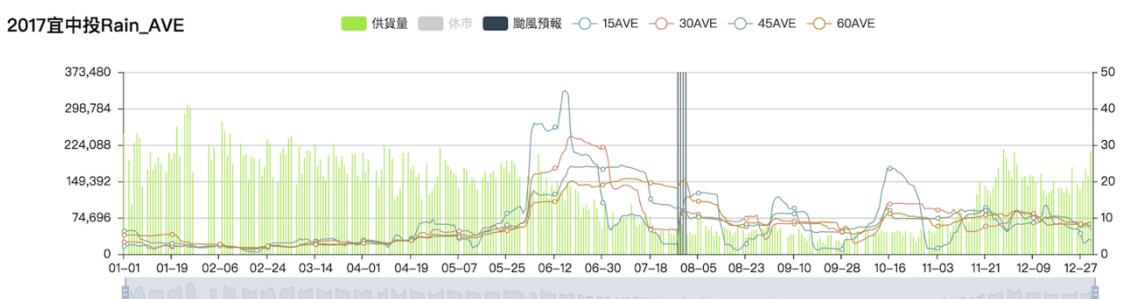
彰化、雲林、嘉義平均溫度分析：



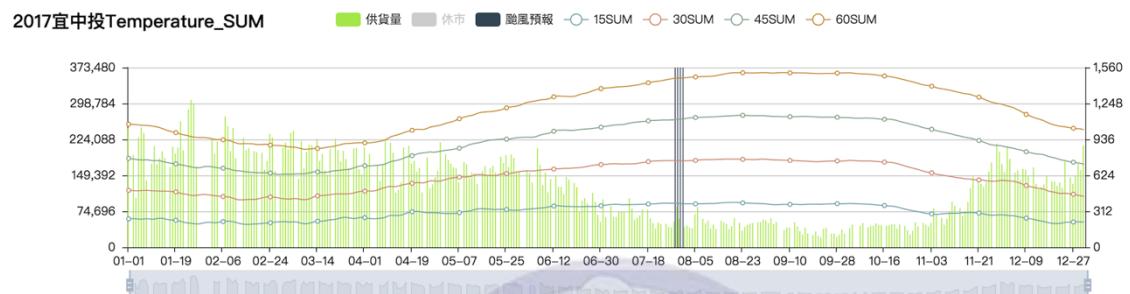
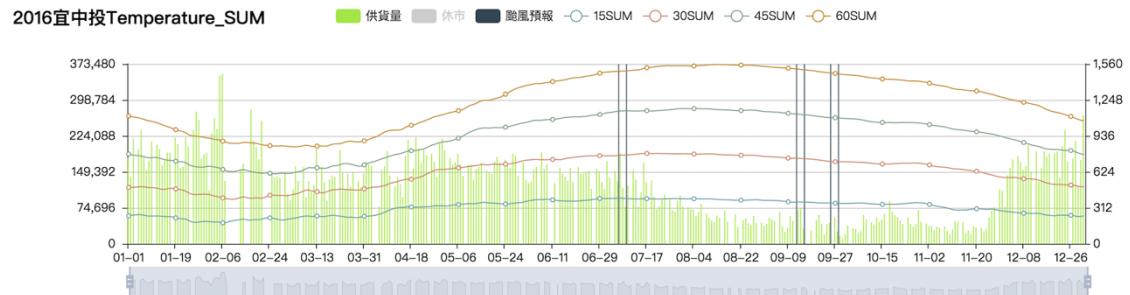
宜蘭、台中、南投累積雨量分析：



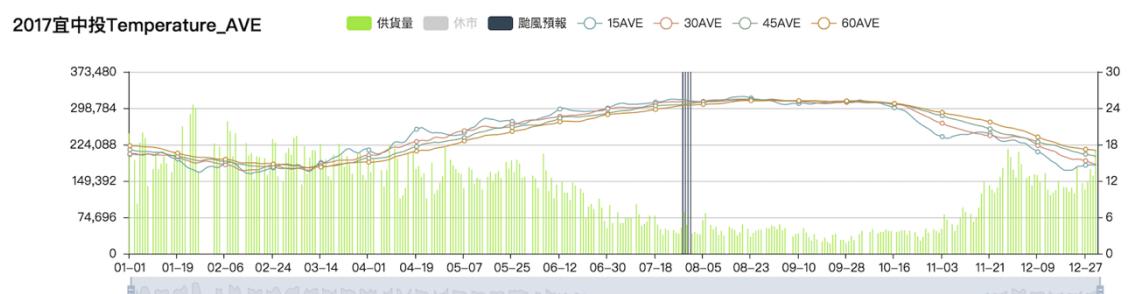
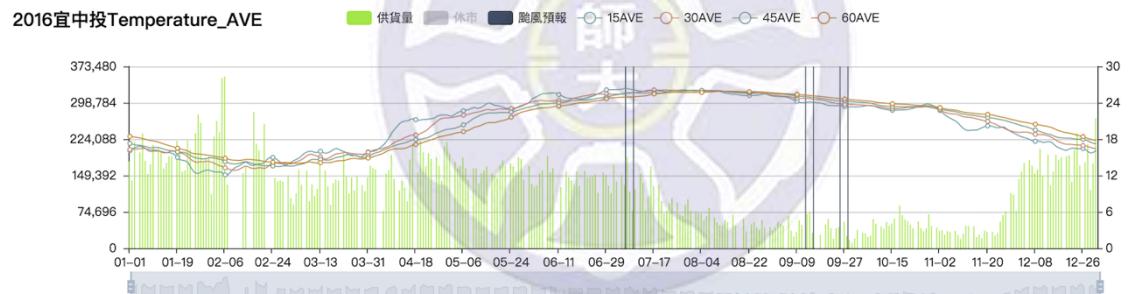
宜蘭、台中、南投平均雨量分析：



宜蘭、台中、南投累積溫度分析：



宜蘭、台中、南投平均溫度分析：



參考文獻

- [1] 黃明德 發行人，1994-09-11，「果菜批發價格是怎麼訂的？」，台東區農業專訊，第 9 期
- [2] 行政院綜合業務處，「行政院第 3322 次院會決議」，101-11-08
- [3] Jason Brownlee, 2014-09-26, “Discover Feature Engineering, How to Engineer Features and How to Get Good at It”
- [4] 机器学习特征工程实用技巧大全：<https://zhuanlan.zhihu.com/p/26444240>
- [5] 高麗菜有這麼多種？口感特色一次看懂！<http://food.ltn.com.tw/article/1243>
- [6] Breiman, L., 1996, “Bagging Predictors”
- [7] Tin Kam Ho, “The Random Subspace Method for Constructing Decision Forests”, 1998-08-18
- [8] Brandon brohrer, “How Recurrent Neural Networks and Long Short-Term Memory Work”
- [9] 古明萱教授，「第五章 作物與溫度」，屏東科技大學農藝學
- [10] 古明萱教授，「第六章 作物與水」，屏東科技大學農藝學
- [11] 農產品交易行情 API 介接說明書：
<https://data.coa.gov.tw/Query/ServiceDetail.aspx?id=037>
- [12] 台北果菜市場 20 天休市 11 天 菜價崩跌：
<https://www.businessweekly.com.tw/article.aspx?id=22017&type=Blog>

[13] 觀測資料說明書：<https://e-service.cwb.gov.tw/HistoryDataQuery/index.jsp>

(開放資料來源)

農產品批發市場交易行情站：

<http://data.coa.gov.tw/Service/OpenData/FromM/FarmTransData.aspx>

中央氣象局測站代號及站況資料查詢：

<https://e-service.cwb.gov.tw/wdps/obs/state.htm>

中央氣象局：觀測資料查詢系統：

<http://e-service.cwb.gov.tw/HistoryDataQuery/index.jsp>

颱風資料庫：有發警報颱風列表：

http://rdc28.cwb.gov.tw/TDB/ntdb/pageControl/ty_warning