

# 運用新聞不確定性與新聞情緒指標預測臺灣加權指數漲跌

報告者：郭東輝

August 26, 2022

# Outline

- ① 研究動機
- ② 建立指標
- ③ 分析方法
- ④ 預測結果
- ⑤ 總結建議

# Outline

1 研究動機

2 建立指標

3 分析方法

4 預測結果

5 總結建議

# 市場情緒

- 現實世界投資人的決策會受到**市場消息**及**個人情緒**的影響，投資人不可能完全保持理性
- 僅透過股價歷史資訊及總體經濟指標來預測股價的方法並沒有考慮到市場消息以及投資人情緒的影響力
- Baker、Bloom and Davis(2016) 三位學者提出一個衡量不確定性的指標-**經濟政策不確定性指標**

# 量化每日臺灣新聞文字的情緒及訊息來增加對臺灣加權指數每日漲跌預測的精準性

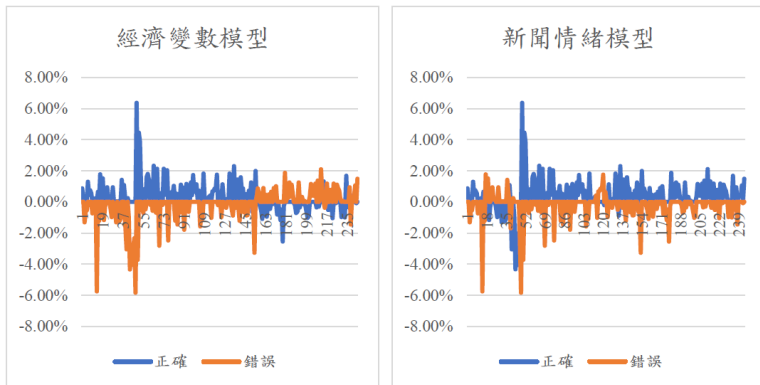


Figure: 股價波動與預測漲跌關係

# Outline

- 1 研究動機
- 2 建立指標
- 3 分析方法
- 4 預測結果
- 5 總結建議

# 經濟政策不確定性指標 (Economic Policy Uncertainty Index，簡稱 EPU 指標)

$$\frac{\text{當月符合EPU關鍵字的新聞篇數}}{\text{當月新聞總篇數}}$$

- Baker、Bloom and Davis(2016) 提出一個 EPU 建構方法分為三個部分：
  - 統計美國「the Wall Street Journal」、「USA Today」等10家大型報社中與經濟政策不確定性有關的文章數目來衡量經濟政策的不確定性
  - 統計每年稅法法條數目的修訂來衡量稅法變動的不確定性
  - 消費者物價指數 (CPI) 和聯邦及地方州的政府支出的不一致來建構
- 原欲參考此公式編制 EPU 指標，惟資料蒐集因當日工商時報的所有新聞篇數無法取得，故斟酌後採由當日符合 EPU 關鍵字的新聞篇數做為衡量臺灣日新聞性的指標。

# 新聞情緒指標

每日情緒指標 = 該目標新聞中所有正向詞次數-該目標新聞中所有負向詞次數

- 新聞情緒指標則是側重在**大眾媒體對景氣和經濟現況的看好與否**
- Baker and Wurgler(2006) 發現當以財經新聞建構之投資人情緒指標很高時，股票對於觀望中的投資人及樂觀主義者會更加有吸引力
- 採字典法進行情緒分析，並採用 NTUSD 提供的中文情感詞典

正向字	不凡的、公平的、友善的、引人注目的、出眾的
負向字	不公平的、不友善的、不正直的、劣等的、不好的



# Google 搜尋趨勢 (Google Trend)

$$G_{day} = g_{month} \times \frac{g_{day}}{\sum g_{day}}$$

- 了解使用者在 Google 上搜尋哪些關鍵字，關鍵字的搜尋熱度會介於 0 到 100 之間，此數值僅代表相對熱度
- Gday 為長期趨勢之日資料，gmonth 為年趨勢之月資料，gday 為月趨勢之日資料，透過轉換之後將每日每個字詞的 gday 做平均，最後獲取每日 Google Trend 指標

# 新聞性指標編制

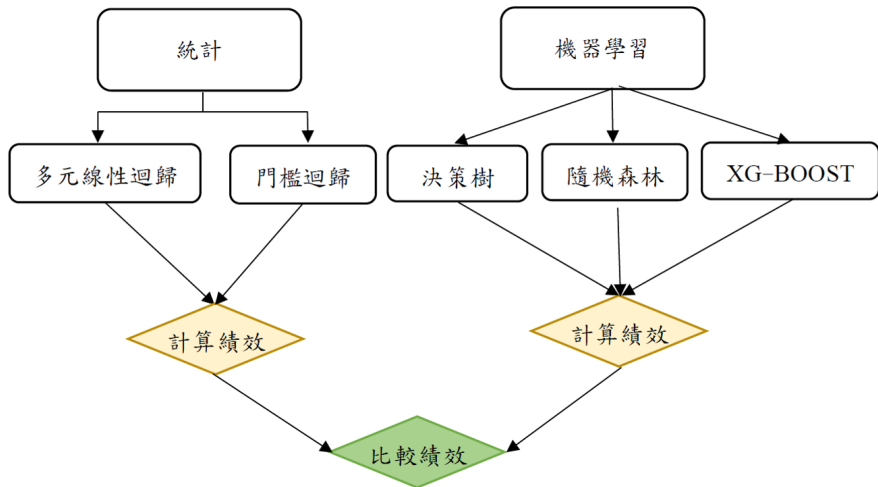
新聞性日指標 = 當日符合關鍵字的新聞篇數來編制新聞性日指標

- 以**工商時報財經新聞**建構的新聞性指標作為衡量臺灣新聞性的標準
- 共整理出 2014/1/2 到 2020/12/31 共 1707 筆資料，並依照使用關鍵字分為**政策性、經濟性、不確定性**三類
- 關鍵字使用：不安，不確定，中央銀行，內閣，公開市場操作……等等（共分為**政策性、經濟性、不確定性**）（關鍵字為自己發想）

# Outline

- 1 研究動機
- 2 建立指標
- 3 分析方法**
- 4 預測結果
- 5 總結建議

# 大綱



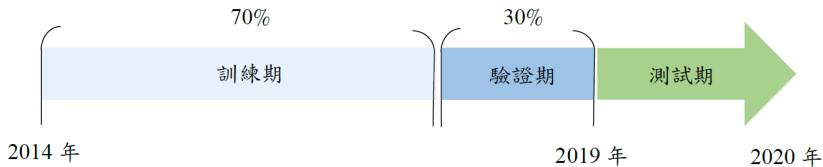
# 模型使用變數表

- **基本變數**(參考文獻): 人民幣兌美元匯率、歐元兌美元匯率、英鎊兌美元匯率、日圓兌美元匯率……等等 (共 17 類)
- **新增變數**: 市場情緒指標、Google 搜尋趨勢 (Google Trend)、新聞性指標編制 (政策性、經濟性、不確定性)
- **模型**使用:
  - 機器學習: 分類樹 (CART 演算法)、隨機森林 (CART 演算法)、Xgboost(AdaBoost 演算法)
  - 統計: 多變量迴歸, 門檻回歸

# Outline

- 1 研究動機
- 2 建立指標
- 3 分析方法
- 4 預測結果**
- 5 總結建議

# 機器學習



- 共使用四組模型：

- 基本變數模型
- 基本變數 + 新聞情緒指標模型
- 基本變數 + Google 搜尋趨勢模型
- 基本變數 + 新聞性指標模型

# 機器學習預測結果

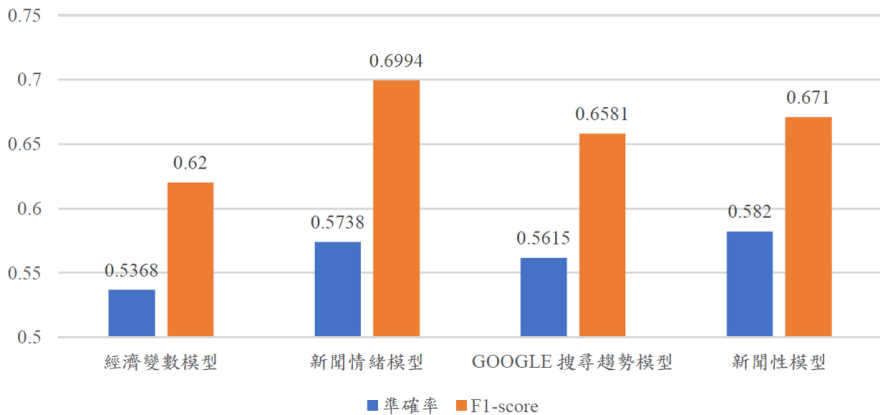
	經濟變數模型	新聞情緒模型	Google 搜尋趨勢模型	新聞性模型
機器學習模型	隨機森林	決策樹	隨機森林	決策樹
準確率	0.5368	0.5738	0.5615	0.582
F1-score	0.62	0.6994	0.6581	0.671

- 從 F1 score 比較，**新聞情緒模型**預測績效最佳
- **Google 搜尋趨勢**在 Google 搜尋趨勢模型和全體變數模型中權重占比都極為重要
- 新增變數對於預測股價漲跌方向是有貢獻的

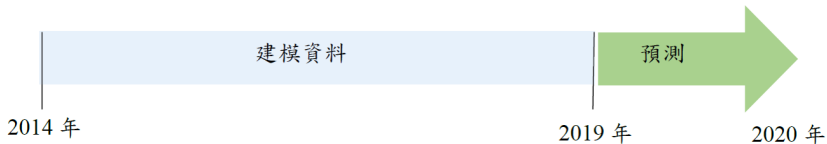


# 機器學習預測結果

## 最適模型比較



# 統計模型

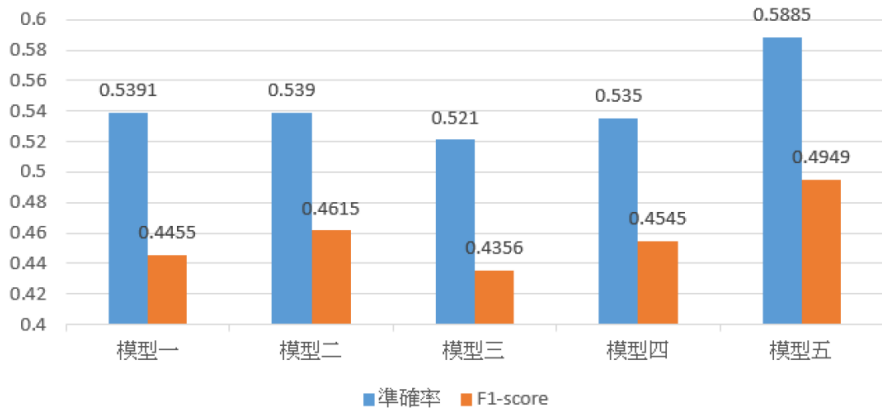


- 共五組模型:

- 基本變數
- 基本變數 + 新聞情緒指標交乘項 (經濟 \* 恐慌指數)
- 基本變數 + Google 搜尋趨勢 \* 恐慌指數
- 基本變數 + 新聞情緒指標交乘項 (不確定 \* 恐慌指數)
- 基本變數 + 模型 1 4 之變數

# 統計模型預測結果

## 最適模型比較



# Outline

- 1 研究動機
- 2 建立指標
- 3 分析方法
- 4 預測結果
- 5 總結建議**

# 結論與建議

- 新增指標皆與臺灣加權指數都不具有線性關係
- 恐慌指數作為交乘項時能卓著改善新增變數顯著影響臺灣加權指數之良窳
- 透過門檻迴歸模型，發現新聞性經濟指標、新聞性政策指標以及Google 搜尋趨勢皆存在門檻值，代表這三項變數在不同期間對於加權指數的影響有所不同
- 機器學習實證結果亦發現在經濟變數加入任一個新增變數皆可提高模型預測能力
- 關鍵字在運用 Google 搜尋趨勢時由於資料時間過短所以不符合預設樣本期間問題
- 偏迴歸係數偏誤問題