# 股票價格預測

資訊碩一李柏漢 | 資訊碩一陳昶安 | 資訊碩一林祐祥 資訊碩一廖偉哲 | 資管碩一林靖淵 | 資管四陳彥融

#### 研究動機

台灣科技產業在全球供應鏈中占有關鍵地位,相關企業的股價 波動不僅受到國內經濟影響,更與國際市場脈動高度連動。對 投資者與研究人員而言,準確預測這些企業的股價變化,能有 效輔助決策、降低風險。然而,台灣科技股具有高波動性與複 雜性,傳統技術分析與經濟模型往往難以掌握其非線性趨勢。 因此,本研究致力於結合資料科學與機器學習技術,探索更具 準確性與前瞻性的股價預測方法。

#### 研究概述

本研究以台灣主要科技類股為分析對象,針對其歷史股價、成交量及技術指標資料進行時間序列建模,採用長短期記憶模型(Long Short-Term Memory, LSTM)進行股價預測。LSTM 為一種具備時間記憶能力的遞迴式神經網路,能有效捕捉股價資料中的長期依賴關係與非線性特徵。透過訓練與測試資料的交叉驗證,評估模型在短期預測上的準確性與穩定性,期望建立一套具實務應用價值的深度學習預測架構,提供投資人更具前瞻性的決策依據。

### 研究流程

探討股市預測的應用價值

爬蟲收集股票資料 整理成CSV檔 處理缺漏值、異常值 正規化股價與技術指標 切分訓練集與測試集

建立 LSTM模型架構 設定參數進行訓練 評估模型準確率 進行比較

探討 預測結果

主題發想

資料搜集

資料前處理

模型建構與訓練

模型評估

\_\_\_

結果分析

資料搜集

本研究採用網路爬蟲(Web Crawling)技術,自TWSE臺灣證 卷交易所網站擷取台灣科技優質股的歷史資料。蒐集項目包含:

- 日收盤價、開高低價、成交量等常見項目
- 技術指標(動能指標、價格變動率等)
- 時間範圍涵蓋近五年資料,以利捕捉長期趨勢

日期	成交股數	成交金額	開盤價	最高價	最低價	收盤價	漲跌價差	成交筆數
110/01/04	9,616,722	2,553,353,170	263.50	271.00	262.50	265.00	2.00	6,878
110/01/05	18,236,900	5,052,768,448	270.00	283.00	269.50	276.50	11.50	12,442
110/01/06	30,972,893	9,123,977,861	284.00	303.50	282.00	298.00	21.50	22,763
110/01/07	25,367,203	7,787,513,900	298.00	321.50	298.00	316.00	18.00	18,991
110/01/08	30,771,661	9,708,435,932	319.00	325.00	306.00	310.00	-6.00	25,031
110/01/11	16,599,358	5,137,076,867	307.00	314.00	302.00	313.00	3.00	13,399

## 資料前處理

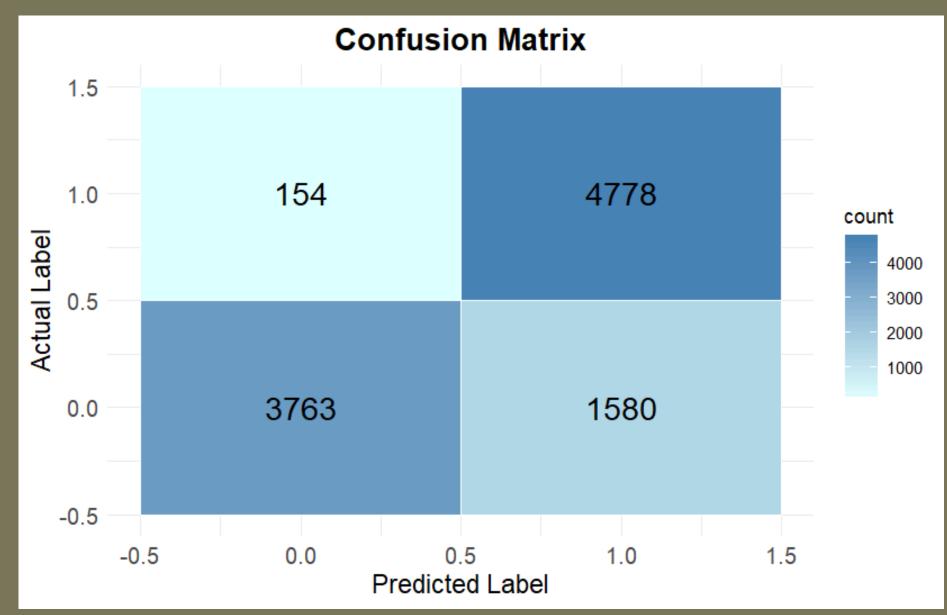
為提升模型預測準確性與穩定性,本研究對原始股價資料進行以下前處理步驟:

- 1. 缺漏值處理
  - 補齊缺漏日期與價格資訊
  - 確保沒有空值
- 2. 技術指標計算
  - 基於原始價格計算動能指標及價格變動率等特徵
  - 每筆資料包含過去多日的技術分析資訊
- 3. 資料標準化
  - 用 Min-Max Scaling 將各欄位特徵壓縮至 [0,1] 範圍
  - 確保不同尺度資料在模型中具同等重要性
- 4. 訓練與測試資料分割
  - 依照時間順序切分訓練集與測試集,避免未來資料洩漏
  - 保持時間序列連貫性,模擬實際預測情境

date	target	open	high	low	close
2021-03-02	1	-0.0546467706738909	-0.0623638107635893	-0.0684959457245138	-0.069970633477983
2021-03-03	0	-0.0664683808750409	-0.0739753510440671	-0.0703510673072648	-0.0681460819398056
2021-03-04	0	-0.0737432179219024	-0.0802277188874014	-0.0805542360123949	-0.0827424942452246
2021-03-05	0	-0.0910209559081984	-0.094518845386451	-0.0889022831347742	-0.0936898034742888
2021-03-08	0	-0.0855648281230523	-0.0936256499802604	-0.0953952086744024	-0.101900285396087
2021-03-09	1	-0.109208048525352	-0.115955535135025	-0.115801546084663	-0.114672146163328

透過這些處理步驟,建立乾淨且具預測意義的特徵資料集,使 LSTM 模型能更有效學習時間依賴性與價格趨勢。

### 結果分析



## 研究結論與未來展望

- 1. 成功以 LSTM 模型預測股價,捕捉時間序列中的非線性與趨勢特徵
- 2. 整合各項技術指標,提升模型在股價預測上的準確率與穩定性
- 3. 從資料蒐集、前處理到模型訓練,形成可複用的預測架構

未來,我們計畫引入更多外部變數,如國際股市指數與新聞情緒,進一步強化模型對市場動態的反應能力。同時,也將探索 Transformer等先進模型架構,提升預測效能與可解釋性。此外, 預計部署自動化爬蟲機制,定期更新資料集,實現即時、智慧的股價 預測平台,協助投資人做出更具前瞻性的決策。