

TSMC 股票預測報告

使用長短期記憶（LSTM）模型進行股市預測的結果分析

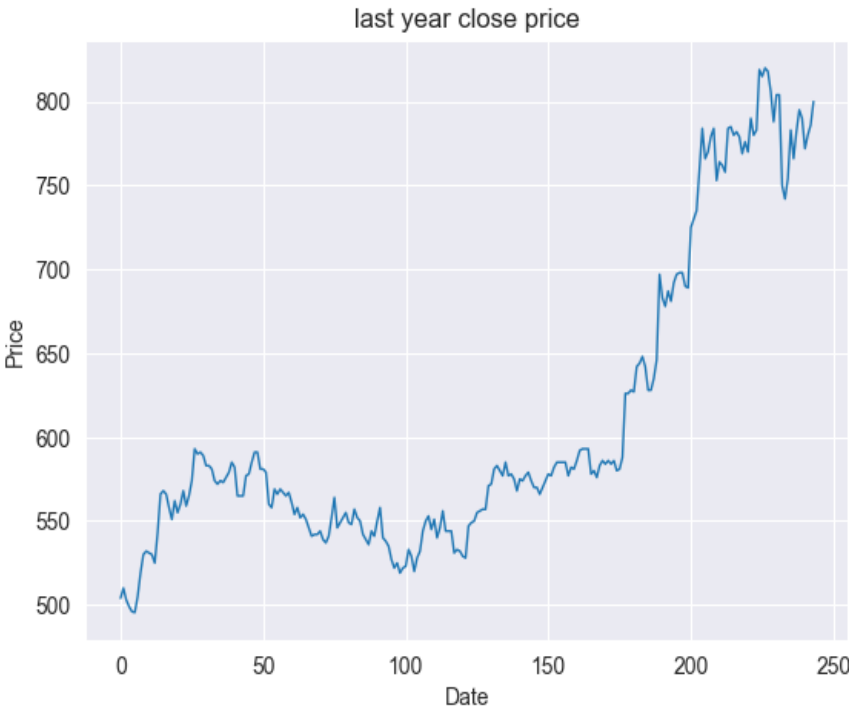
程式	版本	備註
Python	3.9.19	
Pandas	2.2.2	
Numpy	1.23.5	
Scikit-Learn	1.5.0	
tensorflow	2.12.0	
matplotlib	3.9.0	
seaborn	0.13.2	

本報告利用台灣證券交易所提供的過往數據，進行 TSMC 股票價格的預測分析。所用數據集格式為 CSV，包含多個變數如開盤價、收盤價、最高價及最低價。

步驟 1：數據取得與預處理

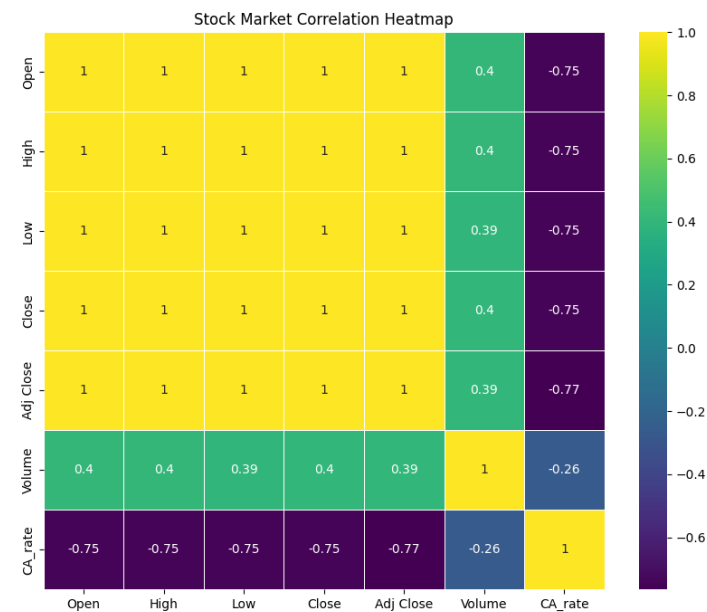
首先，我們從台灣證券交易所獲取 TSMC 過去一年的股票數據（截至 2024 年 5 月 7 日）。隨後，對數據進行可視化處理，繪製出各項數據的時間序列圖，便於觀察趨勢

----->>
這邊輸出的數據就是台積電過往一年的股票曲線，截至我們抓取的當天 (2024-05-07)



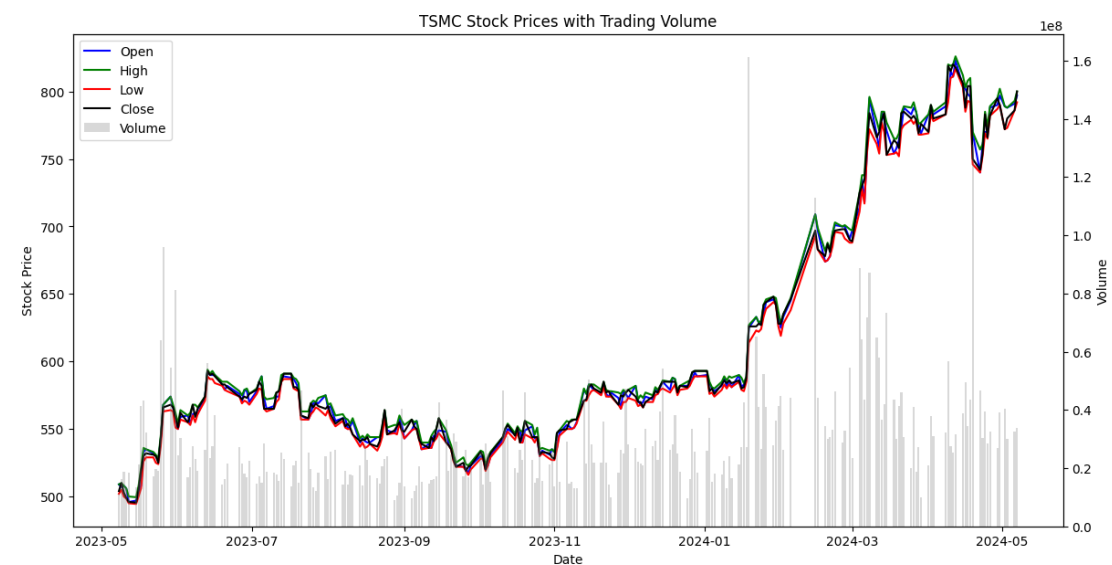
步驟 2：特徵選取

利用熱圖（heatmap）進行特徵相關性分析，顏色越接近黃色表示相關性越高。經分析，開盤價、收盤價、最高價和最低價這四個特徵具有高度相關性，因此選取這四個特徵進行模型訓練。



<< heatmap 會以排列方式快速顯示 2 維資料關聯性，這是我認為相對方便的可視圖

以一年的開盤價、收盤價、最高價和最低價這四個特徵，加上交易人數的棒狀圖以時序輸出



步驟 3：模型構建與訓練

使用 Keras 框架建構 LSTM 模型。在 TensorFlow 2.12.0 版本中，Keras 已內建於 TensorFlow 中。模型結構如下：

- 第一層：包含 128 個神經元的 LSTM 層，輸入形狀為 (35, 128)，即使用 35 天的數據預測下一天的價格。
- 中間層：兩層 LSTM，接著兩層全連接層（Dense layers）。
- 輸出層：單一神經元，預測下一天的收盤價 close。

```
1 from tensorflow.keras.layers import LSTM,Dense
2 from tensorflow.keras.models import Sequential
3
4 model =Sequential()
5 model.add(LSTM(units=128,return_sequences=True,input_shape=(train_x.shape[1],128)))
6 model.add(LSTM(units=50,return_sequences=False))
7 model.add(Dense(units=25))
8 model.add(Dense(units=1))
9
10 model.summary()
```

Python

C:\Users\User\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2k...

super().__init__(**kwargs)

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 35, 128)	66,560
lstm_1 (LSTM)	(None, 50)	35,800
dense (Dense)	(None, 25)	1,275
dense_1 (Dense)	(None, 1)	26

Total params: 103,661 (404.93 KB)

Trainable params: 103,661 (404.93 KB)

步驟 4：模型調整與驗證：

經多次調參，確定了最佳模型參數。為避免過擬合（overfitting），加入了 EarlyStopping 機制。最終模型收斂效果良好。

```

from tensorflow.keras.callbacks import EarlyStopping

callback=EarlyStopping( monitor="val_loss",patience=100,restore_best_weights=True)

model.fit(train_x,train_y,
          validation_split=0.2,
          callbacks=[callback],
          epochs=1000)

```

步驟 5：評估與結果

模型評估採用均方誤差（mean squared error, MSE）指標。模型預測結果的 MSE 值為 2.289771，考慮到預測值在 700 至 800 之間，這一誤差相對較小。結果表明，LSTM 模型在處理時間序列數據方面具有優異的性能。



心得

在進一步放大觀察後，發現模型輸出數據有向右偏移的傾向。為驗證這一點，我進行了純移位計算，結果顯示移位後的 MSE 值有明顯降低。這表明在未來的模型優化中，應考慮對數據進行適當的時間對齊處理，以提高預測準確性。



```
predict = 2.2897713219624696  
shift   = 2.1276107527822123
```

此報告通過實際數據和模型分析，驗證了 LSTM 模型在股票預測中的應用潛力。未來可進一步優化特徵選取和模型參數，提升預測精度。