

模型解釋 (Model Explanation)

本作業使用了兩種主要的分類模型來預測股票的買賣訊號：

1. 羅吉斯迴歸 (Logistic Regression):

- **類型:** 這是一個線性模型，適用於二元分類問題。它透過 Sigmoid 函數將線性組合的輸入特徵轉換為介於 0 到 1 之間的機率，然後根據閾值判斷類別 (0 或 1)。
- **優點:** 簡單、計算效率高、易於解釋。模型的係數可以直接反映各個特徵對預測結果的影響方向和強度。

2. 隨機森林 (Random Forest):

- **類型:** 這是一種基於決策樹的集成學習方法，屬於 Bagging (Bootstrap Aggregating) 的一種。
- **優點:** 能夠捕捉非線性的複雜關係、對過度擬合相對不敏感、對雜訊和離群值具有較好的魯棒性、能夠處理高維度數據並提供特徵重要性排序。

泛化能力評估 (Generalizability Evaluation)

泛化能力是指模型在訓練過程中未見過的新數據上表現良好的能力。本作業通過以下幾個方面來評估模型的泛化能力：

1. K 折交叉驗證 (K-folds Cross Validation):

- 在訓練集上使用了 $K=10$ 的交叉驗證。
- 輸出顯示，Logistic Regression 在交叉驗證中的平均準確率更高 (0.911432 vs 0.914046)，但標準差較大 (0.034850 vs 0.029789)，這表明 Random Forest 在訓練集的不同子集上表現更穩定且更優，預示著其更好的泛化潛力。

2. 評估指標 (Evaluation Metrics):

- **準確率:** 直接衡量整體正確率。測試集上的準確率 (RF 0.89583, LR 0.90625) 提供了模型在未見數據上預測正確率的直接衡量。

3. 回測 (Backtesting):

- 在測試集數據上進行了模擬交易回測。
- 回測是一種將模型預測結果應用於歷史數據來模擬交易策略表現的方法。雖然它不是對未來數據的嚴格泛化，但它提供了模型在實際應用場景（基於歷史數據的模擬）下的性能評估。
- 累積回報圖和累積回報百分比 (RF 511.87%, LR 130.59%) 顯示，Random Forest 策略在測試集上的表現更好。支持 Random Forest 模型在本次分析的數據集上具有更好的實用性和泛化到類似歷史數據上的能力。