**# 7. 模型解釋 (Model Explanation)**

本作業使用了兩種主要的分類模型來預測0050 ETF的買賣訊號：

**## 7.1 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

- 訓練速度快，記憶體需求低

- 對線性可分的數據表現良好

**\*\*在本研究中的表現\*\***:

- 交叉驗證準確率: 0.917 (±0.021)

- 測試集準確率: 0.906

- 最佳參數: C=1000, penalty='l2', solver='saga'

**## 7.2 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

- 不需要特徵標準化

**\*\*在本研究中的表現\*\***:

- 交叉驗證準確率: 0.917 (±0.015)

- 測試集準確率: 0.906

- 最佳參數: bootstrap=False, max\_depth=30, min\_samples\_leaf=1, min\_samples\_split=2, n\_estimators=200

**# 8. 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：

**## 8.1 K 折交叉驗證 (K-folds Cross Validation)**

- **\*\*方法\*\***: 在訓練集上使用了 K=10 的交叉驗證

- **\*\*結果分析\*\***:

  - Logistic Regression: 平均準確率 0.917 (標準差 0.021)

  - Random Forest: 平均準確率 0.917 (標準差 0.015)

  - Random Forest 的標準差較小，表明其在訓練集的不同子集上表現更穩定，預示著更好的泛化潛力

**## 8.2 測試集表現 (Test Set Performance)**

- **\*\*準確率\*\***: 兩個模型在測試集上的準確率相同 (0.906)，直接衡量了模型在未見數據上的預測正確率

- **\*\*混淆矩陣分析\*\***: 兩個模型都展現了良好的分類能力，誤分類情況相似

**## 8.3 交易策略回測 (Trading Strategy Backtesting)**

回測是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：

**\*\*Random Forest 策略表現\*\***:

- 總回報: 237.9%

- 波動度: 0.204

- 夏普比率: 0.126

- 勝率: 53.7%

**\*\*Logistic Regression 策略表現\*\***:

- 總回報: 1143.5%

- 波動度: 0.205

- 夏普比率: 0.159

- 勝率: 60.5%

**\*\*買入持有策略 (基準)\*\***:

- 總回報: 49.1%

- 波動度: 0.291

- 夏普比率: 0.141

**## 8.4 泛化能力結論**

1. **\*\*模型穩定性\*\***: Random Forest 在交叉驗證中展現更好的穩定性（較小的標準差）

2. **\*\*實際應用性能\*\***: Logistic Regression 在交易策略回測中表現更優，累積回報率顯著高於 Random Forest 和買入持有策略

3. **\*\*風險調整回報\*\***: Logistic Regression 的夏普比率(0.159)高於 Random Forest (0.126)和買入持有(0.141)

4. **\*\*策略有效性\*\***: 兩個模型都成功超越了買入持有基準策略，證明了特徵工程和模型預測的有效性

**\*\*最終評估\*\***: 雖然 Random Forest 在模型穩定性方面表現更好，但 Logistic Regression 在實際交易應用中展現了更優的泛化能力和投資回報，這可能是由於金融市場中線性關係的重要性以及羅吉斯迴歸對雜訊的相對魯棒性。