

# 基於機器學習模型對台灣加權指數量化交易研究

作者:朱國安

## 1. 摘要

我們利用 zigzag 指標定義出標籤，使用 23 種技術指標與匯率作為特徵，並利用特徵工程技術，增加差異程度，幫助樹模型，能夠更好的判別方向。

## 2. 前言

金融投資當中有許多因子影響股市，若能使用機器學習，將能幫助我們從高維度的資訊當中，找出方向，協助投資決策。

## 3. 文獻探討

### 3.1 xgboost

xgboost 模型，為梯度提升決策樹(GBDT)中的一種，藉由樹狀結構不斷分支，新的決策樹是根據上一個決策輸出的預測值中，不斷修正，提高模型準確度，該模型可使用於回歸與分類問題當中。

## 4. 實作流程

機器學習中常應用於回歸、分類、分群當中，我們決定使用分類方法，分類出多頭與空頭趨勢，來產生出交易訊號

### 4.1 定義標籤

我們使用 zigzag 指標，產生標籤，定義出過去加權指數多空頭趨勢。Zigzag 指標會根據未來價格波動而改變，定義出區間內的最高點與最低點。該指標可定義參數，繪畫出過去幾天內的最高點或最低點，到目前價位的位置，若過程中回檔大於設定%數，則會固定線圖。

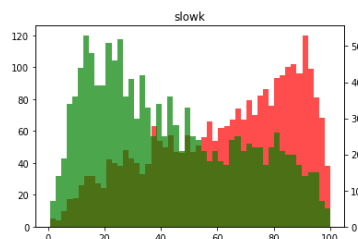


### 4.3 產生特徵

我們使用 ta-lib 套件產生出技術指標，並使用特徵工程技術，增加多空頭時技術指標的差異程度。

使用到的技術指標包含:macd、macdsignal、macdhist、RSI、slowkd、fastkd、sma5、sma5-sma60、收盤價-最大值、willr、ultosc、rocr、apo、sar、vwap5、vwap20、vwap5-vwap20、收盤價-sma5、收盤價-sma20、ForceIndex

並統計加權指數成分股前 100 大中，5、10、20 日均線多頭排列家數，與空頭排列家數，與新台幣匯率每日開收高低資料。



並藉由視覺化篩選特徵值，挑選出差異程度較大的特徵，以上為篩選過後的結果。

### 4.4 模型建立

使用 xgboost 模型，num\_round 設定為 50，訓練集與測試集分別為 80%、20%

## 5. 結論

透過交叉驗證 K-Fold 5，準確率約為 0.88，標準差為 0.017 左右，有良好的分類能力，能夠準確判定多空方向

| in-auc-r ▲ | train-auc-std | st-auc-max | test-auc-std |
|------------|---------------|------------|--------------|
| 0.911632   | 0.00485994    | 0.869489   | 0.0160803    |
| 0.913698   | 0.00460641    | 0.871415   | 0.0151215    |
| 0.915807   | 0.0051563     | 0.873276   | 0.0161191    |
| 0.917824   | 0.00469956    | 0.875737   | 0.0155123    |
| 0.91958    | 0.00499671    | 0.877161   | 0.0157643    |
| 0.921344   | 0.0046278     | 0.878163   | 0.0163662    |
| 0.923184   | 0.00407566    | 0.879495   | 0.0166944    |
| 0.924906   | 0.00448778    | 0.880726   | 0.0170146    |
| 0.926901   | 0.0044651     | 0.882709   | 0.0171534    |
| 0.928361   | 0.00411947    | 0.884195   | 0.0167293    |
| 0.929776   | 0.00369108    | 0.885651   | 0.0170335    |
| 0.93128    | 0.00331454    | 0.887216   | 0.0175155    |
| 0.93287    | 0.00377122    | 0.888676   | 0.017563     |
| 0.934044   | 0.00384986    | 0.88961    | 0.0174951    |
| 0.9358     | 0.00397039    | 0.891384   | 0.0178059    |
| 0.937171   | 0.00415922    | 0.892394   | 0.0175029    |
| 0.938364   | 0.00404716    | 0.893905   | 0.0172182    |
| 0.939805   | 0.00376896    | 0.895097   | 0.0170946    |
| 0.940712   | 0.00353456    | 0.895917   | 0.0172292    |
| 0.941962   | 0.0030259     | 0.897228   | 0.0171542    |
| 0.943333   | 0.00289305    | 0.898281   | 0.0171116    |