台灣ETF價格預測競賽

設計文件

隊伍：TESTTEST

成員：Pan Kuan Lun

# 摘要

股票的預測沒有使用過度複雜的模型與特徵，特徵只有使用漲跌。做特徵選擇後，用 Gradient Boosting訓練預測各etf各天的漲幅。

# 環境

Google Compute Engine

18GB, 2Vcpu

Python + Scikit-learn

# 特徵

國外各式股市指數漲跌（From: <http://finance.yahoo.com/>）、外匯漲跌（From: <http://www.taifex.com.tw/chinese/3/3_5.asp>）、與 ETF 關聯性最高的前100名個股漲跌。取25天。

# 訓練模型

用Boruta做特徵選取， cv 找最佳參數，以 Gradient Boosting 的方式預測各個 ETF 個天的漲跌幅。

# 訓練方式及原始碼

# <https://github.com/jackklpan/tbrain-taiwan-etf-prediction>

# 結論

股票市場由於其複雜性，本身非常難以預測，這次能夠有還可以的成績，運氣絕對是成份之一。也由於股票市場相當複雜且持續變動的特性，若沒有良好的特徵與模型選擇，一個複雜的模型未必會比一個簡單的模型表現來得好。同時，要有機會預測股票市場，比起試驗各種複雜的模型，最重要的是能取得越多不同的有效資訊。但要一次預測到下一週五天的漲跌與價格，如此細節的資訊，我覺得是不太容易達成，相信若改成每天預測隔天的股價漲跌，或是預測較長期的趨勢，會是比較有可能變成不需要靠運氣的比賽。