Machine Learning for Stock Trends: Directional Stock Price Forecast, with Web Applications Available

陳宥任、謝皓雲 、郭承諺、吳秉叡、鄭丞皓、陳聰堯

NCCU
Data Science
2024

INTRODUCTION

Previous studies on stock price momentum have found a positive correlation between the 26-week price divergence rate and future stock returns in the U.S. stock market. Jegadeesh and Titman discovered the presence of momentum effects in the U.S. stock market—by implementing an investment strategy of buying stocks with the best returns over the past 6-12 months and simultaneously selling stocks with the worst returns over the same period, one can achieve economically and statistically significant excess returns. Subsequent research by many scholars has also found that the momentum effect is prevalent across various international stock markets and asset classes.

Additionally, in recent years, regulatory authorities have increased the amount of information disclosure in the stock market and gradually relaxed the limits on stock price movements. These changes have significantly improved the efficiency of the Taiwanese stock market. As a result, the relationship between stock price momentum and expected returns in the Taiwanese stock market is likely to become more pronounced and merits further investigation. Therefore, this study attempts to utilize various technical analysis indicators based on stock prices and trading volumes, combined with historical closing price analyses, to predict future stock market trends.

This study employs various technical analysis indicators to perform machine learning predictions on multiple stocks, aiming to identify stock price patterns in the near future. The goal is to provide investors with actionable investment references on whether to buy or sell specific stocks.

METHODS

After obtaining stock price information from Yahoo Finance, various technical indicators were calculated with the obtained information. For the technical indicators adopted for modelling and their meanings, please refer to Table 1.

Afterward, common ML algorithms such as Random Forest, CART and SVM are applied for two prediction outputs ("Avg3" and "Up3", see Table 2. and Formula 1. and 2. for details), for each prediction output, multiple models were trained. We then selected the best model for each prediction output based on evaluation metrics. In addition, we implemented a dual-model validation mechanism, which would send warning to users if two models produced paradoxical prediction; for example, if model for "Avg3" indicates a rise of the selected stock while model for "Up3" suggests otherwise.

RSI (Relative Strength Index)	Measures the degree of overbuying or overselling of stock prices.
MACD (Moving Average Convergence Divergence)	Indicates the trend and momentum of stock prices through the relationship between two moving averages.
Signal (MACD Signal Line)	The average line of the MACD, used to identify the turning points of the MACD indicator.
CCI (Commodity Channel Index)	Evaluates the variation of stock prices relative to their average stock price, helping users to identify new trends in stock prices.
pctB (%B Indicator)	Describes the relationship between price and Bollinger Bands, used to identify overbought or oversold conditions.
MACD Way	Represents the direction of change in MACD, used to indicate potential short-term price movements.

Table 1. Explanation of the Adopted Technical Indicators

Prediction Output	Explanation	Evaluation Metrics
Avg3 3-Day Average Closing Price (numeric)	The average closing price over the next three trading days.	R-squared
Up3 3-Day Price Change Direction (binary)	The change direction (upward or downward) in price over the next three trading days.	F1-score

Table 2. Explanation of the Prediction Outputs

$$\mathrm{Avg3}_{(t)} = \frac{1}{3}\sum_{n=t+1}^{t+3} \mathrm{Close}_{(n)}$$

$$\mathrm{numUp3}_{(t)} = \frac{\mathrm{Avg3}_{(t)} - \mathrm{Close}_{(t)}}{\mathrm{Close}_{(t)}}$$

$$\mathrm{Up3}_{(t)} = \begin{cases} 1, & \text{if } \mathrm{numUp3}_{(t)} \geq 0 \\ 0, & \text{if } \mathrm{numUp3}_{(t)} < 0 \end{cases}$$

Formula 1. & 2. Formula of the Prediction Outputs

RESULTS

Regression: Due to significant noise and uncontrollable factors in the stock market, accurately predicting closing prices is highly challenging. Models often rely heavily on historical closing prices for predictions, neglecting technical indicators and resulting in a bias towards historical data.

Classification: Technical indicators are inherently suited for predicting price movements in classification models. Historical closing prices, in contrast, do not effectively indicate future price directions. Thus, technical indicators offer superior predictive power for determining stock price directions in classification models.

Table 3. Comparison of "Avg3" Models

Method	Random Forest	Linear Regression	XGBoost
R-squared	0.988	0.978	0.994

Table 4. Comparison of "Up3" Model

Method	RF	CART	SVM	XGB	LogReg
F1-score	0.570	0.524	0.538	0.610	0.542

Figure 1. GUI Showcase

CONCLUSIONS

As mentioned previously, it can be said that modeling on directional output such as "Up3" used in this project is more useful and practical despite the lower accuracy. In the future, we plan to adopt more features in training data such as MA (moving average) from previous days to better capture the time series effects and features like data to better capture the seasonal effects, in order to increase accuracy and benefit more users.

shiny app

github