XGBoost-Based Multi-Factor Stock Selection Model for Rotational Trading

Thanadon Praphutikul
Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand
6470190621@student.chula.ac.th

Yachai Limpiyakorn
Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand
yachai.l@chula.ac.th

ABSTRACT

Unlike traditional buy and sell signals, Rotational trading is a popular method of switching positions between various symbols based on their relative score. Today machine learning techniques are used in various real-world applications, including investments in stock markets. Quantitative investment powered by machine learning would enhance the performance of portfolio formation in financial markets. In this research, we propose an approach of applying a highly efficient gradient boosting tree-based ensemble, XGBoost, for multi-factor stock selection. The two models, monthly and quarterly stock selection were trained on Thailand large-mid capitalization data containing twenty-seven factors which belong to several categories such as value, growth, momentum, liquidity, quality, dividend, and size. It is discovered that the technical factor mainly affects the price movement in monthly XGBoost, whereas the fundamental factor majorly influences the stock changing trends in quarterly XGBoost. The monthly and quarterly rotational portfolio simulation were then performed to evaluate the investing performance measured by portfolio and trade statistics. The three scenarios of monthly and quarterly rotational trading were analyzed. The common findings of the three scenarios reported that the monthly portfolios outperformed in terms of portfolio statistics, due to more opportunities to select new stocks into the portfolio, while in terms of trade statistics, the quarterly portfolios achieved the better results since the longer holding period would reduce noise or whipsaw in trading.

CCS CONCEPTS

• **Computing methodologies** → Machine learning; Machine learning algorithms; Ensemble methods; Boosting.

KEYWORDS

stock selection, multi-factor, rotational portfolio, XGBoost

ACM Reference Format:
Thanadon Praphutikul and Yachai Limpiyakorn. 2023. XGBoost-Based Multi-Factor Stock Selection Model for Rotational Trading. In *2023 5th International Conference on Information Technology and Computer Communications (ITCC 2023), June 15–17, 2023, Tianjin, China.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3606843.3606862

ITCC 2023, June 15–17, 2023, Tianjin, China

ACM ISBN 979-8-4007-0058-3/23/06…$15.00
https://doi.org/10.1145/3606843.3606862

1 INTRODUCTION

One of the key components of investing process is stock selection. However, the stock price was driven by several factors, mainly are fundamental factors and technical factors. Fundamental factors use economics and financial data for business analysis. In 1989, Bernard and Thomas [1] found that the stock price related to the company's growth rate. In other words, the growth factor, categorized as a fundamental factor, will affect the stock selection. The authors described post–earnings-announcement drift phenomenon, or the tendency for a stock's cumulative abnormal returns to drift in the direction of an earnings surprise for several weeks or months following an earnings announcement. Alternatively, technical factors use the price movement of security data for attempting to predict future price movements. In 1993, Jegadeesh and Titman [2] found that the Winner stock or the stocks that in the past had a higher relative performance compared to the others tend to be the future winners, while the Loser stock or those with lower relative performance tend to be future losers. This reflects the effect of momentum factor on stock selection. The momentum factor is categorized as a technical factor. Additional categories of factors that affect the stock selection include value, liquidity, quality, dividend, and size.

In the 1970s, Fama [3] categorized market efficiency into three stages: Strong-form, Semi-Strong-form, and Weak-form. The Strong-form market denotes that investors cannot beat the market since the asset's price already reflects all the information in the market. However, stock prices may not respond to all information promptly. An example of a weak market is the Post-Earnings-Announcement Drift [1], where the stock price gradually rises after the financial statements are announced. Moreover, including the past Winner Stock [2], the asset price continues to rise steadily. Based on these hypotheses, it is evident that the stock market has an anomaly and investors could beat the market with smart portfolio management. This paper thus presents an approach of applying machine learning technology for modeling the complex relationships among various factors that would lead to the objective and strategic investment. The study focuses on the stock data of Thailand large-mid capitalization which are popular among investors and have high liquidity in trading. A tree-based ensemble algorithm, XGBoost (XGB), is selected for learning the complex relationship among the chosen twenty-seven factors categorized in technical and fundamental factors. To reflect the real-world investing, the restriction of trading cost is added when building the rotational portfolios: monthly and quarterly. The performance evaluation of rotational portfolio trading with multi-factor stock selection using

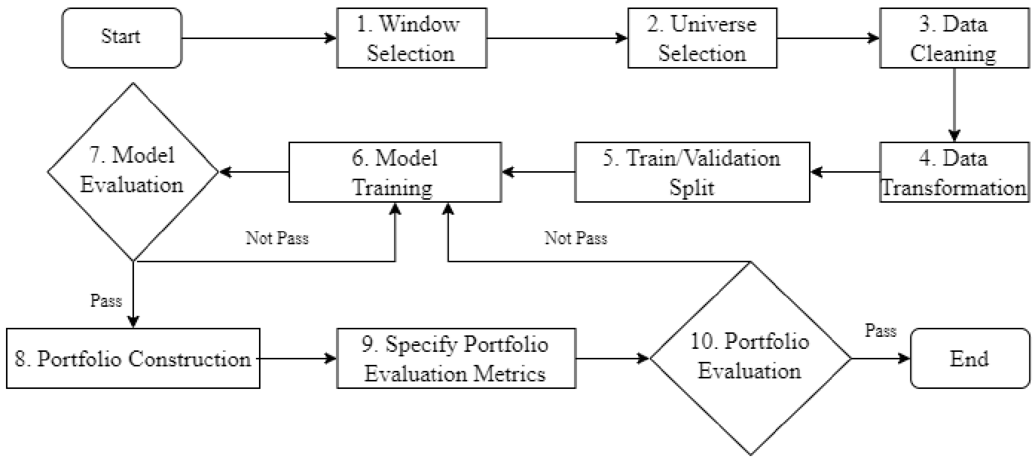

Figure 1: Overview of research methodology.

XGB is compared to the portfolio management using the SET TRI Index and the Equal-Weighted Index widely used for assessing the effectiveness of investment portfolios.

2 BACKGROUND

2.1 Multi-factor Model

A multi-factor model is a financial modeling strategy where multiple factors are used to analyze and explain asset prices [4]. For example, in 1992, Fama and French [5] presented a Three-Factor Model consisting of market risk factors, size factors, and the book value factor to explain expected returns. Later In 1997, the concept was developed to a Four-Factor Model by Carhart [6] and evolved to The Five-Factor Model by Fama and French [7] in 2015. However, these traditional models often describe the relationships between factors as a linear expression. In fact, these factors may not usually be linearly correlated.

Nowadays, machine learning technology has been widely applied to explain these relationships. In 2018, Zhang et al. [8] presented a multi-factor model using Long Short-Term Memory (LSTM) that can detect non-linear correlations. In 2019, Dai and Zhou [9] used Support Vector Regression (SVR) for China A-Shares stock selection based on 500 factors. The result showed that the SVR portfolio yielded an annual return and Sharpe ratio greater than Equal-Weighted (EW). In 2021, Yang [10] presents an approach of applying machine learning algorithms, AdaBoost, for the enhancement of the traditional multi-factor model effect.

2.2 XGBoost

In 2016, Chen and Guestrin [11] invented eXtreme Gradient Boosting (XGBoost), a tree-based ensemble machine learning algorithm with higher predicting power and performance achieved by improvisation on the Gradient Boosting framework by introducing some accurate approximation algorithms. The applications of XGBoost have been applied in many fields. In 2019, Hsieh et al. [12] used the XGBoost to detect the human body's response to stress. In 2020, Li et al. [13] found that XGBoost was effective in feature selection and classification for assessing credit problems on large databases.

In the financial area, Li and Zhang [14] proposed a multi-factor stock selection strategy in 2018. The XGBoost model was trained to learn dynamic weighting of seven factors. The result achieved an annual return of 22.54% that outperformed the Equal-Weighted, and the Information coefficient weighting strategies. In 2019, Zhange and Chen [15] developed a multi-factor stock selection model based on XGBoost, Logistic Regression (LR), and Support Vector Machines (SVM). The experiment reported that the portfolio advised by XGB yielded 134% of return which outperformed the benchmark of 28%. The XGBoost also provided the better result compared to the other two algorithms of stock classification. In 2021, Zhong et al. [16] constructed various multi-factor models using Logistic Regression (LR), XGBoost, and Linear Regression. The result showed that the LR and XGB provided higher returns compared to the Linear Regression model.

3 METHODOLOGY

Figure 1 illustrates the steps of methodology presented in this paper. The overall process is described in the following subsections.

3.1 Window Selection

The dataset contains the ten years of end-of-day historical data from January 2012 to December 2021, provided by an investing research and service company. The data collected from year 2012 to 2019 are used for model construction, while the data collected from 2020 to 2021 are used for portfolio construction and evaluation. The proportion of the size of data used for model construction and portfolio construction is 80 : 20. Figure 2 illustrates the movement of the SET TRI Index, which is a composite index that represents the price movement for all common stocks trading on the SET (The Stock Exchange of Thailand). Observing that the period of 2020 to 2021 had covered all bullish and bearish markets. Due to the COVID-19 pandemic outbreak at the beginning of 2020, the SET TRI Index had collapsed around -34.46% since 2019, and then showed a rising trend at the end of 2020 onwards.

Figure 2: SET TRI Index movement during the period of year 2012 to 2021.

3.2 Universe Selection

The top 200 stocks with the highest market capitalization at the end of 2019 were selected to represent Thailand's large and mid-capitalization stocks. To prevent the lookahead bias and the survivorship bias, any stock delisted from the markets during the period of 2020 to 2021 were not excluded from our universe. Due to the findings of Baquero in 2005 [17], these biases could lead to an unrealistic good result.

3.3 Data Cleaning

The data was arranged in a cross-section. Normally, missing data and outliers can be found in the dataset. In this work, the null value was replaced with zero. For example, when the data is naturally less than one year from the trading day, it will disable the calculation of the one-year rate-of-change that requires a minimum of 250 data points. Here the null value was then replaced with zero to enable the metric computation. Outlier data are handled by leaving as is since outliers can regularly happen in financial data. In literature, Mandelbrot 1967 [18] found that large changes tend to be continued by large changes, and small changes tend to be followed by small changes, also known as volatility clustering. In 2020, Chong et al. [19] found that the unusual trading volumes could predict abnormal returns close to the earnings announcement date.

3.4 Data Transformation

The input format for model learning contains the columns of stock acronym, date, a list of twenty-seven factors, and class label. Each factor value was transformed to ranking number, and then transformed to decile (0-9) or divided into ten groups based on its percentile order. The excerpt of 10 stocks out of the total number of 100 stocks is shown in Figure 3. The data in the first row inform that on 2020-01-02, the value of factor PE was ranked to the number 1-10. Next, the value of RankPE was further transformed to the value 0-9 depends on its percentile order ranging 0-9. Here, the total number of example stocks is 100, the stocks with factor ranking 1-10 will be categorized in the first group with the Decile value of 0. Finally, the class label of each instance was assigned to three classes: 0 (Bottom30%), 1 (Middle40%), and 2 (Top30%), based on one-month returns (20 trading days) and three-months or quarterly returns (60 trading days).

3.5 Train/Validation Split

The dataset prepared for model construction (period of January 2012 to December 2019) was divided into 80% of train set and 20% of validation set. The train set was used to fit the model while the validation set was used for gauging model performance and fine-tuning the model hyperparameters.

3.6 Model Training

The two models, monthly and quarterly stock selection, were trained with XGB algorithm for study in this work. The dataset contains twenty-seven features chosen from technical factors and fundamental factors by the expert. The predicted output is the probability of a stock falling into three distinct classes, which are Bottom30%, Middle40%, and Top30%.

3.7 Model Evaluation

The model performance is measured by ROC curve (Receiver Operating Characteristic curve) and AUC (Area under the ROC Curve).

3.8 Portfolio Construction

The investment portfolio is built to verify the effectiveness of stock selection strategy suggested by the XGBoost model. For each of the two models, monthly and quarterly rotational trading systems were built with back-testing simulation using the test set collected during January 2020 to December 2021. The rule of selecting the stocks to the portfolio is picking the top tenth percentile (20 stocks) with the highest probability of each predicted class. The position sizing of all stocks in the portfolio is set to equal weighted and the executions price is set as At-The-Open Price (ATO).

In the real world, there are costs of buying and selling marketable securities, such as commission and slippage. Commission, also

Symbol	Date/Time	PE	RankPE	Decile RankPE	Decile RankPBV	...	20D Returns	20D Returns Class	60D Returns	60D Returns Class
AWC	2020-01-02	215.74	1	0	9	...	-5.26	1	-12.74	2
CKP	2020-01-02	107.52	2	0	4	...	-18	0	-40	1
CIMBT	2020-01-02	91.64	3	0	9	...	26.92	2	-21.02	2
GULF	2020-01-02	79.5	4	0	0	...	15.68	2	-46.21	0
GPSC	2020-01-02	71.48	5	0	0	...	2.61	2	-3.05	2
VGI	2020-01-02	68.46	6	0	1	...	-10.66	0	-29.35	1
BGRIM	2020-01-02	67.34	7	0	1	...	21.23	2	-37.23	1
ACE	2020-01-02	63.65	8	0	9	...	-30.23	0	-37.5	1
TRUE	2020-01-02	63.27	9	0	6	...	-16.96	0	-40.4	0
THG	2020-01-02	57.11	10	0	2	...	0	1	-32.47	1
...							...			

Figure 3: Example of Data Transformation.

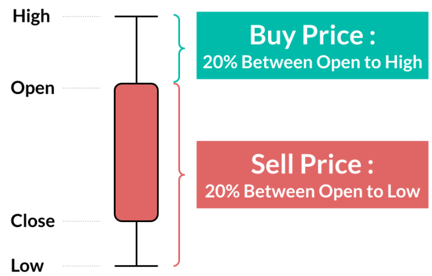

Figure 4: Example of volatility slippage calculation for buy price and sell price.

known as trading fee, is the cost that brokers have charged an investor for using their services to execute the orders [20]. The slippage is the difference between the expected price of a trade which in this case is the open price and the price at which the trade is executed [21]. In this paper, the commission fee is set to 0.15% per trade (total 0.3% for buy and sell), whereas the slippage is set to 20% as volatility slippage threshold per trade. The volatility slippage of buy price and sell price are computed as (1) and (2), respectively. Figure 4 illustrates the calculation of volatility slippage, of which the value is dynamic based on stock daily trading range. Observing that the stock with high volatility will have a wide range resulting in a higher slippage value. While the stock with low volatility will have a lower slippage value.

$$Buy\ Price = Open\ + (\ slippage\ threshold * (\ High - Open\ )) \quad (1)$$

$$Sell\ Price = Open\ - (\ slippage\ threshold * (\ Open - Low\ )) \quad (2)$$

3.9 Specify Portfolio Evaluation Metrics

Trade metrics and portfolio metrics are used for investing result assessment. Trade metrics consist of number of trades (# Trades), percentage of profitable trades (% Win), percentage of losing trades (% Loss), average trade profit and loss or Trade Expectancy, and average bar holding (Avg. Bar Held) is calculated by the average period from buy to sell of the entire portfolio. The portfolio metrics consist of three dimensions of portfolio statistics as described in the following subsections.

3.9.1 Return. Considering annual portfolio return, also known as the Compound Annual Growth Rate (CAGR).

3.9.2 Risk. In terms of risk, this paper uses 2 metrics which are: 1) the annualized standard deviation (Ann. STD) of the entire portfolio's return to represent the portfolio volatility, and 2) the portfolio maximum drawdown (Max. DD) to represent the maximum percentage that investors could lose during investing period, referring optimization trading strategies of Pardo [22].

3.9.3 Risk-Adjusted Return. In addition to risk measurement described in the previous subsection, the risk-adjusted return is also measured with two metrics: Sharpe ratio and MAR Ratio. The Sharpe ratio was proposed by William F. Sharpe 1994 [23], calculated by portfolio excess returns from risk-free rate divided by portfolio standard deviation, whereas the MAR Ratio is calculated by CAGR divided by Max. DD.

3.10 Portfolio Evaluation

The return of investment from portfolio created by the XGBoost stock selection strategy is compared to the benchmark consisting of the SET TRI Index, and Equal-Weighted Index. The Stock Exchange of Thailand Total Return Index (SET TRI) is a market-capitalization weighted index, which is calculated by the Stock Exchange of Thailand. The index is an industry standard used to evaluate portfolio performance that represents all types of return on security investment including dividends, capital gains, and right issues [24]. Since the SET TRI Index is market-capitalization weighted, i.e., the stock with higher market capitalization has more impact on the index movement. In other words, this index could not represent the entire portfolio performance. The Equal-Weighted index is thus an alternative that provides the identical importance to individual stock in a portfolio. That is, the stocks with the smallest market capitalization in universe are given equal statistical significance, or weight, to the largest market capitalization companies [25]. In this work, building the Equal-Weighted portfolio initially selected all stocks in the universe and rebalanced position size at the beginning of the month to prevent any single stock from having too much influence on the portfolio. Without adjusting position size throughout the test period, the stocks with high price growth will gain greater

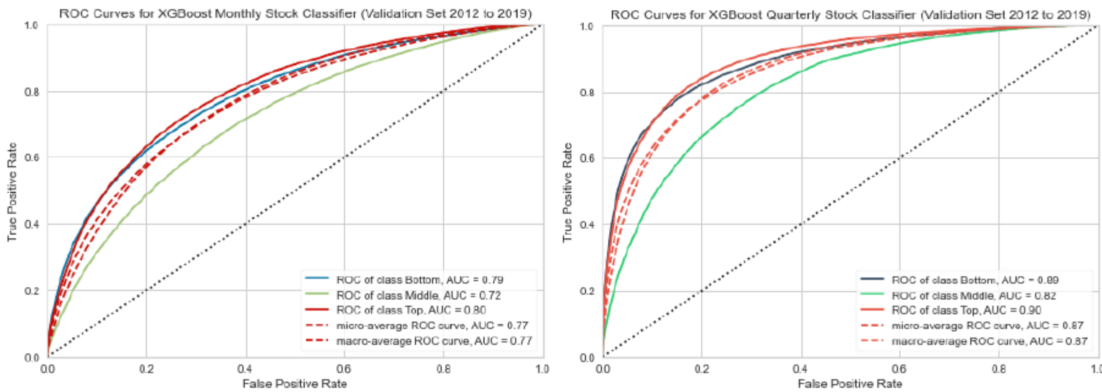

Figure 5: ROC-AUC of monthly stock selection model (left) and quarterly stock selection model (right) on validation set.

weight than others, resulting in the portfolio that fails to represent all the stock movements.

4 RESULTS AND DISCUSSION

4.1 Model Performance Evaluation

During model training, grid search with 3-fold cross-validation was applied for tuning the hyperparameters. It was observed that the *max_depth* was the main influence parameter, and the optimal value was set to 8 for training the monthly and quarterly stock selection models. Figure 5 shows the ROC-AUC on the validation set during training the monthly stock selection model (left) and the quarterly stock selection model (right). The AUC reports the results of classifying the monthly/ quarterly returns of the Top, Middle, and Bottom Classes. All the AUC of quarterly stock selection model are higher than that of monthly stock selection model. Compared to the monthly model, the curves of quarterly stock-selection increase faster and closer to the upper-bound. The prediction of the stock returns in longer period, quarterly, is better than the monthly returns.

A benefit of XGBoost is that the model provides a built-in function to rank features ordered by their importance. The model relies more on the features with the higher score for prediction. Figure 6 (left) shows the important factors for monthly stock selection. The first two factors consist of the momentum factors including 52 Weeks High (H52W) and past stock rate-of-change 250 days (ROC250); followed by the growth factor including rolling twelve-month revenue growth based on Quarterly-on-Quarterly (RevenueQoQ); size factor as stock market capitalization (MKC); intermediate-term momentum as past stock rate-of-change 60 days (ROC60) and very long-term momentum as stock all time high score (ATH). For monthly stock selection, the momentum factors were mainly used for analyzing over multiple periods to predict the price trends.

As shown in Figure 6 (right), the fundamental factor plays an important role in the quarterly stock selection. The first three ranking features include rolling twelve-month revenue growth based on Year-on-Year (RevenueYoY), rolling twelve-month earnings before interest and taxes based on Quarter-on-Quarter (EBITQoQ), and rolling twelve-month revenue growth based on Quarter-on-Quarter (RevenueQoQ). These features are categorized as the growth factor which belongs to the fundamental factor. Next ranking is value factor as debt ratio, followed by size factor as stock market capitalization (MKC), and momentum factor as 52 Weeks High (H52W) and All Time High (ATH). It is evident that the quarterly stock price movement tends to depend on the fundamental factor rather than the technical factor.

As shown in Figure 7, the performance of XGBoost monthly and quarterly models has declined, compared to the performance on the validation sets. During 2020-2021, the period of test data collected, the stock market was more volatile due to the COVID-19 pandemic, causing stock market regime changed. However, the measures still provided good discrimination during the period of COVID-19 crisis.

4.2 Portfolio Comparisons

Based on the constructed XGBoost models, monthly and quarterly, the investment portfolios were created using the test set. The backtesting simulation was carried out to build monthly and quarterly rotational trading system from January 2020 to December 2021. Those stocks at the top tenth percentile with the highest probability of prediction or *Top-Class stocks* were selected to the portfolio. In this work, 1) the cumulative returns of monthly and quarterly XGBoost portfolios are compared to study the discrepancies between the two models. 2) study the effects of adding the trading cost to compute the cumulative returns, and 3) compare the performance of XGBoost portfolios with the benchmark namely the SETTRI index and Equal-weighted.

Firstly, comparison of cumulative returns between the XGBoost monthly and quarterly portfolios was carried out. Figure 8 visualizes the Cumulative returns of XGBoost stock selection monthly and quarterly. Both monthly and quarterly rotational trading provided the constant growth or increasing trends of investing returns, except for the first quarter of year 2020 that portfolio returns had declined due to COVID-19 pandemic that resulted in the decreased

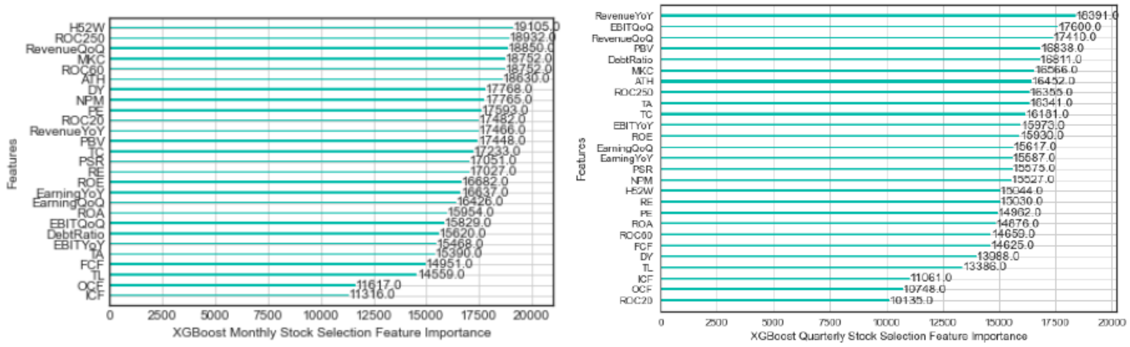

Figure 6: Ranking of feature importance of monthly stock selection model (left) and quarterly stock selection model (right).

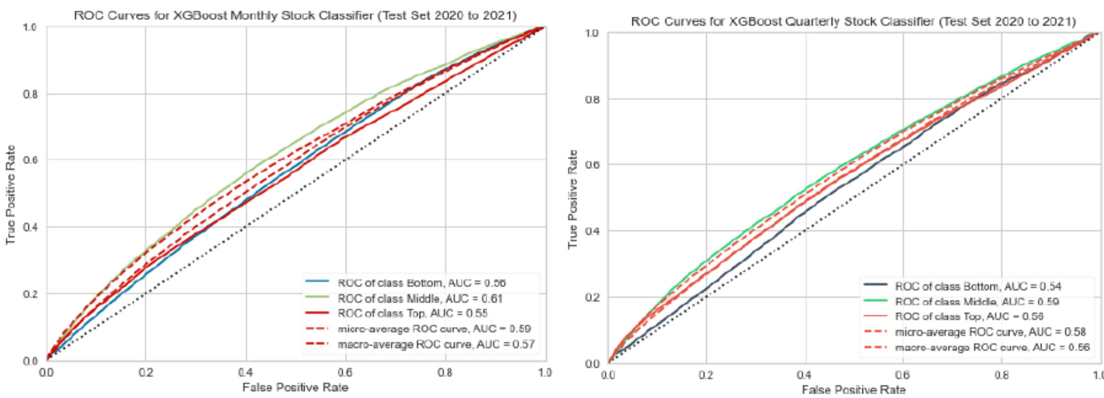

Figure 7: ROC-AUC of monthly stock selection model (left) and quarterly stock selection model (right) on test set.

values of monthly and quarterly Max. Drawdown to -37.16% and -42.18% respectively as shown in Table 1. The annual portfolio return (CAGR) of the Top-class using XGBoost monthly stock selection gained the higher return of 48.19% compared to 30.96% of the quarterly stock selection. The monthly stock selection could mitigate the risk by reallocating the stocks in the portfolio every month. Therefore, the monthly portfolio has more chance to select the best predicted stocks, while the quarterly portfolio must wait for a quarter to select new stocks into the portfolio. Finally, the risk-adjusted return including The Sharpe Ratio and MAR Ratio of the monthly portfolio also gain the better performance compared to the quarterly portfolio.

In terms of trade statistics, the monthly portfolio contains a large number of trades which is 439 trades during the test period, compared to the quarterly portfolio containing 149 trades. The percentage of wining trades and losing trades of each condition have similar profile. But the average trades profit/loss or trade expectancy of quarterly portfolio is significantly better than monthly portfolio since the longer holding period can easily achieve much more profit compared to the shorter holding period. The Avg. Bar Held of 20.57 (monthly) and 61.85 (quarterly) confirm the average period from buy to sell of the entire portfolios, that is, 20 days and 60 days for monthly and quarterly rotational trading.

Secondly, to reflect the real-world stock investment, the trade commission of 0.15% and volatility slippage of 20% are added to the entire portfolios. Table 2 reports that with added trading cost, the XGBoost monthly stock selection annualized return (CAGR) drastically decreased from 48.19% (without trading costs) to 29.26%, while the quarterly stock selection annualized return slightly declined from 30.96% (without trading costs) to 27.26%. It is assumed that the more frequently of trading, the more trading cost will affect the trade expectancy. As shown in Table 2, the monthly portfolio trade expectancy decreased from 3.78% (without trading costs) to 2.64%, while the quarterly portfolio trade expectancy (8.82%) is close to the former (9.49%).

Figure 9 visualizes the Cumulative returns of XGBoost stock selection monthly and quarterly. Similar to Figure 8, both monthly and quarterly rotational trading provided the constant growth of investing returns, except for the first quarter of year 2020 that both portfolio returns had declined due to COVID-19 pandemic. Assuming that the more frequently of trading, the more trading cost incurs that lessen the investment return. It is observed that the

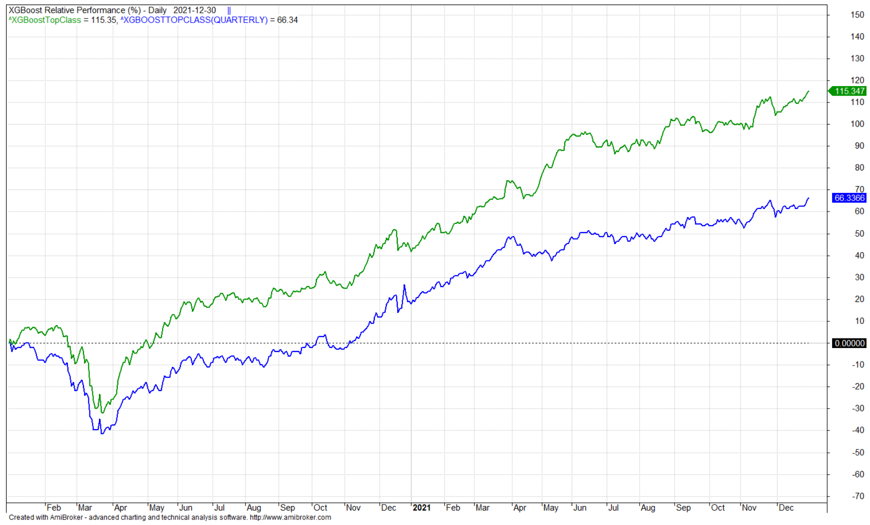

Figure 8: Cumulative return of XGBoost stock selection monthly (green line) and quarterly (blue line).

Table 1: Comparisons of Portfolio statistics of XGBoost stock selection monthly and quarterly

Portfolio statistics	Monthly Top-Class	Quarterly Top-Class
CAGR (%)	**48.19**	30.96
Max. Drawdown (%)	**-37.16**	-42.17
Ann. STD (%)	**25.16**	25.68
Sharpe Ratio	**1.76**	1.05
MAR Ratio	**1.3**	0.73
Trade statistics	Monthly Top-Class	Quarterly Top-Class
#Trades	439	149
% Win	52.16	**54.36**
% Loss	47.84	**45.64**
Trade Expectancy %	3.78	**9.49**
Avg. Bar Held	20.57	61.85

Table 2: Comparisons of Portfolio statistics of XGBoost stock selection monthly and quarterly with trading cost

Portfolio statistics	Monthly Top-Class	Quarterly Top-Class
CAGR (%)	**29.26**	27.26
Max. Drawdown (%)	**-37.15**	-40.46
Ann. STD (%)	24.02	**24.78**
Sharpe Ratio	**1.05**	0.94
MAR Ratio	**0.79**	0.67
Trade statistics	Monthly Top-Class	Quarterly Top-Class
#Trades	424	144
% Win	49.53	**54.17**
% Loss	50.47	**45.83**
Trade Expectancy %	2.64	**8.82**
Avg. Bar Held	20.59	61.91

gap between the two curves is narrower compared to the curves in Figure 8.

Lastly, the performance of XGBoost portfolios with trading cost are compared with the benchmarks: SET TRI Index and Equal-Weighted Index, as illustrated in Figure 10. Table 3 summarizes

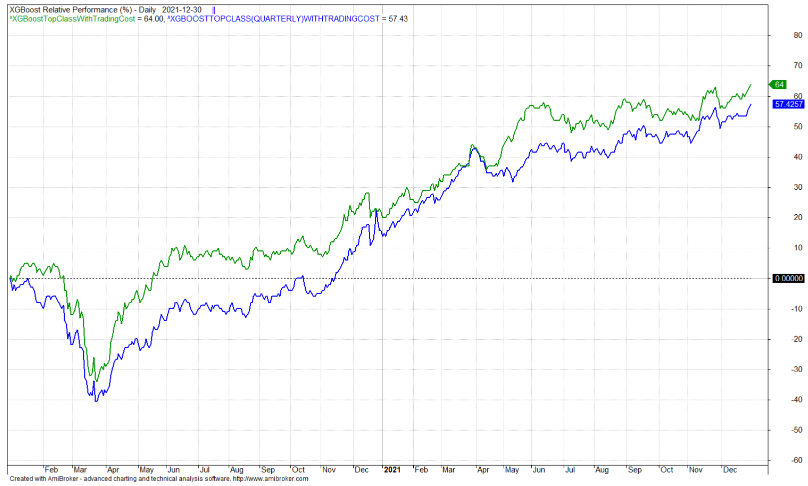

Figure 9: Cumulative return with trading cost of XGBoost stock selection monthly (green line) and quarterly (blue line).

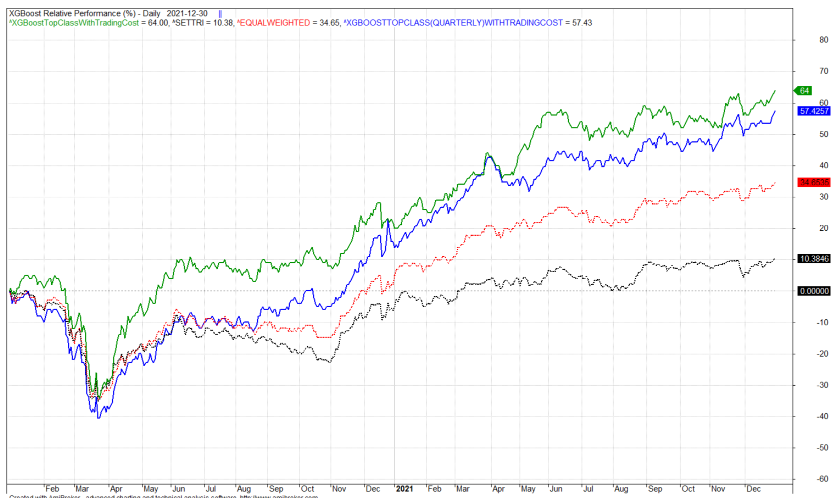

Figure 10: Cumulative return of XGBoost of Top-Class monthly (green line) and quarterly (blue line) stock selection with trading cost, SET TRI Index (black dash line) and Equal-Weighted (red dash line).

the portfolio statistics of XGBoost, SET TRI, and Equal-weighted. The results showed that the values of CAGR of SET TRI and Equal-weighted during the test period are 5.28% and 16.82%, while the monthly and quarterly portfolios with Top-class stocks advised by XGBoost yielded the better returns of 29.26% and 27.26% respectively. It is evident that the returns of XGBoost significantly beat the market.

In terms of risk, during the COVID-19 pandemic, the Max. drawdown of XGB monthly portfolio is -37.15%, closing to that of SET TRI Index and Equal-Weighted, while the Max. drawdown of XGB quarterly portfolio is -40.46% which indicates the higher risk compared to the two benchmarks. The values of annualized standard deviation of the XGB monthly and quarterly portfolios are slightly higher than that of the two benchmarks since the portfolio volatility is calculated based on both moving up and down of portfolio. In terms of risk-adjusted return defined as the Sharpe ratio and MAR ratio, the XGBoost Top-class quarterly and monthly stock selection, again, achieve the higher values compared to that of SET TRI Index and Equal-Weighted Index.

5 CONCLUSION

This paper presents an approach of multi-factor stock selection using a highly efficient gradient-boosting decision trees, XGBoost. The study focuses on rotational trading including monthly and quarterly. The two XGBoost models were trained on the dataset of Thailand Large-Mid Capitalization to learn the complex relationships among twenty-seven features for classifying the stocks into Top30%, Middle40%, and Bottom30% classes. The findings report that the quarterly stock price movement tends to depend on the fundamental factor, contrary to the monthly stock selection that the momentum indicators, which belong to the technical factor,

Table 3: Portfolio statistics of top class XGBoost monthly and quarterly stock selection with trading cost, SET TRI Index and Equal-Weighted

Portfolio statistics	Monthly Top-Class	Quarterly Top-Class	Equal-Weighted	SET TRI Index
CAGR (%)	**29.26**	27.26	16.82	5.28
Max. Drawdown (%)	-37.15	-40.46	**-34.9**	-35.31
Ann. STD (%)	24.02	24.78	**19.95**	22.81
Sharpe Ratio	**1.05**	0.94	0.64	0.06
MAR Ratio	**0.79**	0.67	0.48	0.15
Trade statistics	Monthly Top-Class	Quarterly Top-Class	Equal-Weighted	SET TRI Index
#Trades	424	144	4,613	N/A
% Win	49.53	**54.17**	49.32	N/A
% Loss	50.47	**45.83**	50.68	N/A
Trade Expectancy %	2.64	**8.82**	1.59	N/A
Avg. Bar Held	20.59	61.91	20.44	N/A

are mainly used to predict the price trends. The monthly and quarterly rotational portfolios were built to evaluate the effectiveness of the stocks selected by the models. The back-testing simulations were carried out to evaluate the portfolio performance in three scenarios: 1) comparison of the performance of the monthly and quarterly portfolios, 2) comparison of the performance of the rotational portfolios with added trading cost, and 3) comparison of the performance of the rotational portfolios with the benchmarks. The results report that all the cumulative returns of XGBoost rotational portfolios provides the constant growth and realistic drop during the Covid-19 pandemic. The common findings of the three scenarios reported that in terms of portfolio statistics, the monthly portfolios achieved the better value in all dimensions because there are more opportunities to select new stocks into the portfolio. However, in terms of trade statistics, the quarterly portfolios achieved the better results as the longer holding period would reduce noise or whipsaw in trading.

ACKNOWLEDGMENTS

The dataset used in this research is supported by SiamQuant Co., Ltd. which is an investing research and service company in Thailand. The stock price data had been adjusted to the later corporate dividend payouts, stock splits, or the issue of additional shares had been considered. And the timestamp was marked on the fundamental data based on financial announcement days to prevent a lookahead bias.

REFERENCES

[1] Victor L. Bernard, Jacob K. Thomas. 1989. Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?. Journal of Accounting Research, Vol. 27, Current Studies on The Information Content of Accounting Earnings (1989), pp. 1-36

[2] Narasimhan Jegadeesh and Sheridan Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance Vol. 48, No. 1 (Mar. 1993), pp. 65-91

[3] Eugene F. Fama. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, May 1970, Vol. 25, No. 2, Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York, N.Y. December, 28-30, 1969 (May 1970), pp. 383-417

[4] Investopedia. Multi-Factor Model: Definition and Formula for Comparing Factors Retrieved 11 Febuary, 2023 from https://www.investopedia.com/terms/m/multifactor-model.asp

[5] Eugene F. Fama and Kenneth R. French. 1992. The cross-section of expected stock returns. The Journal of Finance, vol. 47, no. 2, pp. 427-465

[6] Mark M. Carhart. 1997. On Persistence in Mutual Fund Performance. The Journal of Finance. Mar 1997, Vol. 52, No. 1, pp. 57-82

[7] Eugene F. Fama and Kenneth R. French. 2015. A five-factor asset pricing model. Journal of Financial Economics, 116(1), pp. 1-22

[8] Ru Zhang, Chenyu Huang, Weijian Zhang, and Shaozhen Chen. 2018. Multi Factor Stock Selection Model Based on LSTM. International Journal of Economics and Finance, Vol. 10, No. 8

[9] Dai Jingyi and Zhou Jingwei. 2019. A Novel Quantitative Stock Selection Model Based on Support Vector Regression. 2019 International Conference on Economics Management and Model Engineering (ICEMME)

[10] Ming Yang. 2021. Research on quantitative multi-factor stock selection model based on machine learning. 2021 the 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST)

[11] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkkd international conference on knowledge discovery and data mining, pp. 785-794

[12] Cheng-Ping Hsieh, Yi-Ta Chen, Win-ken Beh, and An-Yue A. Wu. 2019. Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection. IEEE International Workshop on Signal Processing Systems (SiPS), 2019.

[13] Hua Li, Yumeng Cao, Siwen Li, Jianbin Zhao, and Yutong Sun. 2020. XGBoost Model and Its Application to Personal Credit Evaluation. IEEE Intelligent Systems, Vol 35, Issue 3, pp. 52-61

[14] Jidong Li and Ran Zhang. 2018. Dynamic Weighting Multi Factor Stock Selection Strategy Based on XGboost Machine Learning. 2018 IEEE International Conference of Safety Produce Informatization (IICSPI)

[15] Xiaoyun Zhange and Wanyi Chen. 2019. Stock Selection based on Extreme Gradient Boosting. 2019 the 38th Chinese Control Conference (CCC2019), July 27-30, 2019, Guangzhou, China

[16] Yihua Zhong, Lan Luo, Xinyi Wang, and Jinlian Yang. 2021. Multi-factor Stock Selection Model Based on Machine Learning. Engineering Letters, Volume 29, Issue 1

[17] Guillermo Baquero, Jenke ter Horst, and Marno Verbeek. 2005. Survival, Look-Ahead Bias, and Persistence in Hedge Fund Performance. Journal of Financial and Quantitative Analysis, Vol. 40, No. 3, September 2005

[18] Benoit Mandelbrot. 1967. The variation of certain speculative prices. Journal of Business, Vol. 40, No. 4, pp. 393–413

[19] Terence T. L. Chong, Yueer Wu, and Jue Su. 2020. The Unusual Trading Volume and Earnings Surprises in China's Market. Journal of Risk and Financial Management

[20] Investopedia. How Do I Keep Commissions and Fess Eating Trading Profits?. Retrieved 09 April, 2023 from https://www.investopedia.com/ask/answers/08/trading-frequency-commissions.asp

[21] Investopedia. Slippage: What It Means in Finance. Retrieved 09 April, 2023 from https://www.investopedia.com/terms/s/slippage.asp

[22] Robert E. Pardo. 2008. The Evaluation and Optimization Trading Strategies, Second Edition, page 265

[23] William F. Sharpe. 1994. The Sharpe Ratio. The Journal of Portfolio Management Fall 1994, 21 (1), pp. 49-58, doi: https://doi.org/10.3905/jpm.1994.409501