Industrial Engineering and Applications
L.-C. Tang (Ed.)

doi:10.3233/ATDE230054

Analyzing Impact of Economic Indicators on Vietnam Stock Market Using Machine Learning Techniques

Nuttawan Sangsawai[a] and Daricha Sutivong[a]
[a] *Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand*

Abstract. Many people find interest in stock markets because of the potential financial gains. Understanding the fundamentals of each stock market is crucial as each has its own unique traits and driving factors. Traditionally, statistical methods are commonly used to find the relationship between various economic indicators and the stock markets. This study aims to utilize a different approach, namely machine learning techniques, a widely used tool for data analytics, to analyze the impact of economic indicators on the Vietnam stock index, which is a rising market during the past decade. The investigated machine learning algorithms include tree-based algorithms such as Decision Tree, Random Forest, and XGBoost. Monthly data, totaling 257 observations from August 2000 to December 2021, were used in this study. The results reveal that the XGBoost algorithm achieves the highest accuracy at 96.67% and the five most influential variables affecting the Vietnam stock market are S&P 500 index, consumer price index, exports, imports, and oil price, respectively, all with a positive relationship, while the relationships of the exchange rate, unemployment rate, and GDP with the Vietnam stock market are unclear.

Keywords. machine learning, Vietnam, stocks, economic indicators, data analytics

1. Introduction

Over the last decade, Vietnam's economy has emerged as one of ASEAN's rising stars, with an average gross domestic product (GDP) growth of around 6–7%, and being one of the few countries to achieve positive economic growth during the COVID-19 pandemic in 2020. Furthermore, the Vietnam government has also projected that by 2035, the country will have an upper middle-income status and a GDP of 1 trillion USD. Due to the country's rapid economic growth and investment promotion policies, the Vietnam stock market is now one of the most attractive markets in the world, with the VN Index growing by 36% in 2021 as shown in Figure 1. Vietnam's stock market is still regarded as a frontier market with limited market size, despite its rapid growth since its founding in 2000 [1].

According to [2], the researcher examined the relationship between stock prices and macroeconomic indicators in the US and found that macroeconomic factors significantly affect the stock prices. Additionally, numerous researchers have explored the same subject and other macroeconomic indicators in other countries, but majority of the studies were carried out in developed countries [3] using various statistical techniques.

In recent years, due to the enhanced capability of computers and technology, machine learning algorithms have received increasing popularity. Many studies show that machine learning algorithms outperform traditional statistical models in terms of accuracy in stock price prediction [4]. Moreover, other factors, such as commodity prices and foreign economies, may have an impact on stock markets.

Therefore, the main purpose of this study is to investigate the factors, such as domestic and global indicators, that may have an impact on the Vietnam stock market. Models are developed using various machine learning techniques and their performance will be compared. We will then utilize analyses of important factors within these techniques to explain the impact of these indicators on the stock market.

Figure 1. VN Index from established to 2021

2. Literature Review

Many studies have examined how macroeconomic indicators affect the stock market. However, the results are still inconclusive. For instance, a high inflation leads the rising price of goods and thus reduces purchasing power. As a result, it affects average people's living expenses and may also reduce company earnings due to higher input costs. Hence, an inflation is expected to have a negative impact on stock prices. However, past work reveals conflicting results. While [2] found a negative relationship,[5] reported a positive relationship between the inflation rate and the U.S. stock markets.

Besides the U.S., other markets have also been examined. [6] studied the effects of macroeconomic and microeconomic factors on company stock prices in the Czech Republic from 2006 to 2016 using the Johansen and Juselius cointegration test, the Hansen cointegration test, and the VECM model to examine the long run relationships between selected indicators and stock prices. The research results show that industrial production, gross domestic product, and profitability ratios significantly influence stock prices in the long run. [7] shows that interest rates and inflation significantly influence stock prices in Ghana by using Johansen's cointegration test and the VECM model. The OLS method was employed by [8] to determine how macroeconomic factors affect various portfolios in the Istanbul stock market. They discovered that inflation has positive effects on all portfolios, but only seven out of thirteen of these effects are significant. They also found that the exchange rate has a variety of effects and the unemployment rate has positive effects, but none of them are significant, and that only two of the ten portfolios are differently significantly impacted by the money supply.

In Asia, the long run and short run relationships between macroeconomic factors and the Malaysian stock market indices for the years 1977 to 2011 were examined by [9] using the ARDL bounds test approach. The study demonstrated that, over the long run, GDP has a positive impact on stock indices, while exchange rates, inflation, money supply, and a producer price index have a negative impact. However, in the short term, inflation and the money supply have a positive impact, while exchange rates, GDP, and the producer price index have a negative impact. Regarding the relationship between exchange rates and the stock prices, [10] found that in eight Asian countries, namely Hong Kong, Indonesia, Japan, Korea, Malaysia, Philippines, Singapore, and Thailand, the exchange rates and stock prices have only a short run relationship from 1991 to 2005.

In Vietnam, many studies have been conducted using various approaches and involve different economic indicators. [11] examined the impact of macroeconomic factors on stock prices in the short and long run through The Vietnam Ho Chi Minh Stock Index (VN Index), using the ARDL model. The research results show that in the long run, there's no relationship between world oil prices and interest rates on the VN Index, while the money supply has a positive effect, and the exchange rate has a negative effect on the stock market. However, both interest rates and exchange rates have a negative effect in the short term. [12] used the ECM model and the Granger causality test to examine the long term and short term relationships between stock price and macroeconomic variables. In the long term, the oil price, money supply, and an interest rate have a positive impact on the VN Index, while the exchange rate and SJC gold price have a negative impact, and the consumer price index does not show a significant impact. In the short term, the index has a positive relationship with oil price volatility and a negative relationship with interest rate variability. The Granger causality test also reveals that oil prices, money supply, and interest rates have a causal relationship with the VN Index. Using the VECM model, [13] studied the impact of six macroeconomic factors on the VN Index, including the consumer price index (CPI), industrial production indices, interest rates, the VND/USD exchange rate, retail oil prices, and gold prices. The study shows that the VN Index has a positive relationship with industrial production indices and retail oil prices. The rest of the indicators, namely the consumer price index, interest rates, exchange rate, and gold prices, negatively affect the VN Index. According to [14], the money supply has a positive relationship with the VN Index. On the other hand, the index had a negative relationship with exchange rates and inflation.

Other than economic indicators, numerous research works have reported that the Vietnam stock market has been influenced by global economies and commodity prices. Macroeconomic news updates from the U.S. had a substantial impact on the Vietnam stock market, as demonstrated by [15]. According to [3], there is a strong positive correlation between Vietnam stock prices and the US industrial sector and the US money market, as well as a positive relationship between VN Index and the S&P 500 [3,16]. Regarding the relationship between commodity prices and the Vietnam stock market, [16,17] found that Vietnam stock prices have a positive relationship with the oil price while the VN Index shows a negative relationship with the gold price.

Instead of traditional statistical methods as widely used above, the study employs machine learning techniques to examine how economic indicators affect the Vietnam stock market. These techniques not only demonstrate the effects of the indicators similar to conventional techniques but also quantify the indicators' relative importance using feature importance techniques, which will be described in section III. Since in the real world many factors are normally involved, ranking the importance of these factors can contribute significantly to an understanding of the situation.

3. Background

3.1. Decision Tree

A decision tree is a supervised learning algorithm that can be used for both classification and regression. It creates a tree-like structure by dividing the data into branches; each represents a possible decision or an outcome and that demonstrates how one decision leads to the next and illustrates that each option is mutually exclusive. The tree's root node denotes the start of the process, followed by decision nodes, denoting the division of the data, and leaf nodes, denoting the outcomes. This splitting process begins at the root node and continues until a leaf node is reached. It is impossible to further separate the leaf node.

Decision Tree for regression is also known as regression tree. A regression tree is used to predict continuous valued outputs rather than discrete outputs. Classification And Regression Trees (CART) employs least squares in regression scenarios, where splits are logically chosen to reduce the residual sum of squares between the observation and the mean in each node. In mathematics, RSS (residual sum of squares) can be expressed as shown in Eq. (1).

$$\mathrm{RSS} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{1}$$

In which, n is the number of observations, Y_i is actual values, and $\hat{Y}_i$ is predicted values.

3.2. Random Forest

A supervised learning approach called a random forest is created by individually creating each decision tree and then combining the trees using a bagging method or a bootstrap aggregation. The random forest uses a majority vote for classification and an average value for regression at the end of the procedure rather than relying on just one decision tree. The number of trees needs to be optimized for better accuracy while avoiding overfitting.

3.3. XGBoost

Gradient Boosting Decision Trees (GBDTs) is a decision tree ensemble learning approach similar to the random forest. However, XGBoost technique uses gradient boosting to iteratively combine many decision trees into a stronger learner. The objective loss function is optimized from the residual at each iteration to increase the previous predictor's prediction accuracy. To avoid overfitting, a regularization term is also included in the objective function.

3.4. Performance Comparison

3.4.1. Coefficient of Determination or R-Squared (R^2)

The coefficient of determination is an important statistical indicator of regression models that examines how well a model predicts or explains the outcomes. The more trustworthy the model, the higher the value. The calculation is shown in Eq. (2).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \overline{Y}_i)^2} \tag{2}$$

In which, n is the number of observations, Y_i is actual values, $\hat{Y}_i$ is predicted values and $\overline{Y}_i$ is an average of actual values.

3.4.2. Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is one of the indicators to measure the errors of the model. Its value is the square root of the average of the squares of the errors, the difference between the predicted values and the actual value. A lower value indicates more accuracy. The calculation is shown in Eq. (3).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} \tag{3}$$

In which, n is the number of observations, Y_i is actual values, and $\hat{Y}_i$ is predicted values.

3.5. Pearson Correlation Coefficient

The linear relationship between two sets of data is measured using the Pearson correlation coefficient, often known as Pearson's r. Its values may be between -1 and 1. Values of 1 and -1 indicate a direct and flawlessly positive and negative relationship, respectively. No linear relationship exists when the correlation coefficient is 0. The calculation is shown in Eq. (4).

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{4}$$

In which, cov(X,Y) is the covariance of variables X and Y, σ_x is the standard deviation of X, and σ_Y is the standard deviation of Y.

3.6. Feature Importance

3.6.1. Feature importance

Feature importance is a method to score each input feature for a certain model indicating the importance of each feature. A higher score means that a specific attribute affects the model more significantly. There are various techniques for calculating feature importance, but the most commonly used are Gini Importance and Permutation Feature Importance. Gini Importance calculates feature importance by a reduction in a node's impurity weighted by the node probability, i.e., the number of samples that reach that

node from the total number of samples. Permutation Feature Importance calculates feature importance by calculating the change in error as a result of permuting a feature's values. The characteristic is crucial to the model if permuting the values significantly alters the inaccuracy.

3.6.2. SHAP value (SHapley Additive exPlanations)

The SHAP value is one of the feature importance techniques. It determines the contribution based on cooperative game theory by supposing that each feature is a player in a game where the prediction output is the payout. Players who contribute to a game more than others may receive a greater reward. Shapley shows that a fair payout among the players depends on their contribution to the game.

4. Research methodology

4.1. Data

The data used in this study is monthly data for the period August 2000 to December 2021, a total of 257 observations. The VN Index, S&P 500 index, Brent oil price, and exchange rate (Vietnam Dongs to US Dollars) come from the Investing.com website, and other economic indicators, such as the consumer price index (which measures inflation), imports, exports, GDP, and the unemployment rate, come from the tradingeconomics.com website. Note that the unemployment rate and GDP values are a quarterly and annual data, respectively. The unemployment rate and GDP value are assumed to be constant for all months within a given quarter or a year.

4.2. Methodology

In this study, we use machine learning techniques to examine how economic indicators affect the Vietnam stock market. We employ three different tree-based machine learning models, namely decision tree, random forest, and XGBoost. Besides being applicable to both classification and regression problems, these tree-based techniques provide many advantages. They can be used on data with nonlinear relationship and without data preparation and pre-processing such as data normalization and feature scaling. Moreover, they do not have assumptions on relationship between features. They are also easy to interpret and visualize, and their algorithms are fast compared to others such as neural networks.

We divide the data into training (80%) and test (20%) sets. We then randomly divide the training data into ten groups of equal size to undertake the 10-fold cross validation. The associated R-squared and RMSE are calculated. The performance of the model is evaluated by applying the hyperparameters with the highest cross validation accuracy to the test set.

5. Results and Discussion

5.1. Pearson Correlation Coefficients

The Pearson correlation coefficients are shown in Table 1. The CPI, the unemployment rate, the exchange rate, and Brent oil price are four of the eight features that have a low-to-medium linear relationship with the stock price, specifically with correlation coefficients of 0.705, 0.238, -0.640 and 0.220, respectively. Additionally, there are strong linear correlations among these eight features. These characteristics illustrate that the conventional multiple linear regression may be inappropriate for analyzing the VN Index.

Besides the exchange rate, all features exhibit a positive linear relationship with the stock price, according to the signs of the correlation coefficients.

Table 1. Pearson correlation coefficients of factors and VN index

	Price	**Total Exports**	**Total Imports**	**CPI**	**Unem**	**GDP**	**VND/ USD**	**Brent**	**S&P 500**
Price	1								
Total Exports	0.806	1							
Total Imports	0.821	0.994	1						
CPI	0.705	0.923	0.929	1					
Unem	0.238	0.177	0.195	0.089	1				
GDP	0.763	0.962	0.965	0.985	0.133	1			
VND/ USD	-0.640	-0.871	-0.879	-0.984	-0.096	-0.956	1		
Brent	0.220	0.161	0.195	0.365	-0.026	0.267	-0.395	1	
S&P500	0.853	0.942	0.940	0.803	0.259	0.878	-0.719	0.040	1

5.2. Performance Analysis

The R-square and RMSE of decision trees, random forests, and XGBoost are displayed in Table 2 for the performance analysis. The XGBoost algorithm yields the most accurate model, with the lowest RMSE of 56.07 and the highest R-squared of 96.67%. The decision tree approach is the second-best model, with an R-squared of 91.85% and RMSE of 87.71. The random forest is last with an R-squared of 90.01% and RMSE of 97.09.

Table 2. Performance and accuracies of three models

Model	**R-squared**	**RMSE**
Decision Tree	0.9185	87.71
Random Forest	0.9001	97.09
XGBoost	0.9667	56.07

5.3. Feature Importance

We use the SHAP value to investigate the feature importance in order to analyze which indicators in the dataset have stronger effects on the Vietnam stock index. Since the XGBoost model is the most accurate model, its SHAP is utilized and displayed in Figure 2.

According to the feature importance from Figure. 2, the S&P 500 index has the most impact on the model, which implies it has the greatest impact on the VN Index when compared to other features. The following features include consumer price index (CPI), the total exports, Brent oil price, the total imports, and the exchange rate (VND/USD), respectively. The unemployment rate and GDP are seen to have relatively small impact on VN Index.

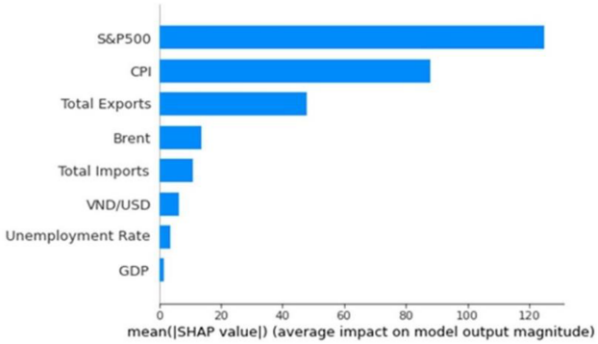

Figure 2. Feature importance based on the SHAP value

Figure. 3 displays the feature contribution from the SHAP value. Note that higher values of the S&P 500 index, consumer price index (CPI), the total exports, Brent oil price, and the total imports have a greater positive impact on the model, or the pink dots representing high values of the features are on the right side signifying a positive impact on VN Index. Therefore, there exists an obvious positive association between the S&P 500 index, consumer price index (CPI), the total exports, Brent oil price, and the total imports and VN Index. However, the exchange rate (VND/USD), the unemployment rate, and GDP do not seem to have a clear directional association with VN Index, as low and high (blue and pink dots) of the features lead to both positive and negative impact on VN Index.

Additionally, the contributions of the S&P 500 index, the consumer price index (CPI), and the price of Brent oil are clearly asymmetrical. High indices of the S&P 500 and the CPI have greater effects than low indices, while low prices of Brent oil have higher effects than high prices.

Note that our results are similar to and different from previous research in certain aspects. Specifically, in previous research, the S&P 500 has a positive relationship with Vietnam stock prices [3], while the exchange rate and consumer price index (CPI) have a negative relationship, and the oil price has a positive relationship [14]. However, [10] found no relationship between the exchange rate and stock prices in eight Asian countries.

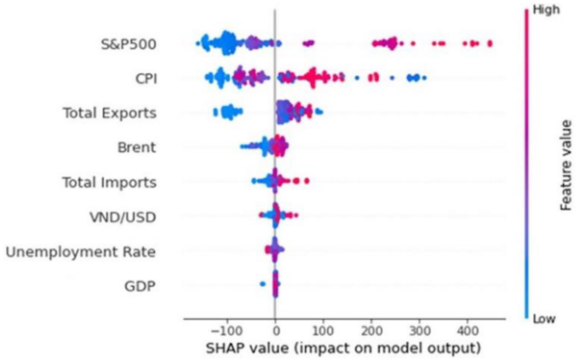

Figure 3. Feature contribution based on the SHAP value

6. Conclusions

Many data analysis techniques can be used to examine the effects of economic indicators. The majority of earlier research works commonly employ various statistical regression techniques to accomplish this goal and only examine the effects of certain selected indicators. In this study, we use three machine learning methods, namely decision trees, random forest, and XGBoost to analyze the effects of economic variables on the Vietnam stock market. In terms of both R-square and RMSE, we discover that the XGBoost algorithm offers the most accurate prediction. Additionally, analysis of feature importance by SHAP value reveals that the S&P500 index, consumer price index (CPI), the total exports, Brent oil price, and the total imports all have a favorable influence on VN Index in respective order. The effects of the GDP, unemployment rate, and exchange rate (VND/USD) are not as obvious. Future research may include additional elements, such as political indicators, or explore additional methods, such as statistical regression and other advanced machine learning approaches.

References

[1] Nguyen AP, Nguyen HA, Ho THM, Ngo PT, McMillan D. Risk and returns of different foreign ownership portfolios: Evidence from Vietnam stock market. Cogent Economics & Finance. 2019;7(1).
[2] Fama EF. Stock Returns, Real Activity, Inflation, and Money. The American Economic Review. 1981;71(4):545-65.
[3] Hussainey K, Khanh Ngoc L. The impact of macroeconomic indicators on Vietnamese stock prices. The Journal of Risk Finance. 2009;10(4):321-32.
[4] Bhattacharjee I, Bhattacharja P. Stock Price Prediction: A Comparative Study between Traditional Statistical Approach and Machine Learning Approach. 2019 4th International Conference on Electrical Information and Communication Technology (EICT)2019. p. 1-6.
[5] Boudoukh J, Richardson M. Stock Returns and Inflation: A Long-Horizon Perspective. The American Economic Review. 1993;83(5):1346-55.
[6] Pražák T. The Effect of Economic Factors on Performance of the Stock Market in the Czech Republic. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis. 2018;66(6):1613-26.
[7] Adam A, Tweneboah G. Macroeconomic Factors and Stock Market Movement: Evidence from Ghana. SSRN Electronic Journal. 2008.
[8] Rjoub H, Türsoy T, Günsel N. The effects of macroeconomic factors on stock returns: Istanbul Stock Market. Studies in Economics and Finance. 2009;26(1):36-45.

[9] Bekhet HA, Mugableh MI. Investigating Equilibrium Relationship between Macroeconomic Variables and Malaysian Stock Market Index through Bounds Tests Approach. International Journal of Economics and Finance. 2012;4(10).
[10] Narayan P, Smyth R, Lean HH. Exchange Rate and Stock Price Interaction in Major Asian Markets: Evidence for Individual Countries and Panels Allowing for Structural Breaks. The Singapore Economic Review (SER). 2012;56:255-77.
[11] DAO HT, VU LH, PHAM TL, NGUYEN KT. Macro-Economic Factors Affecting the Vietnam Stock Price Index: An Application of the ARDL Model. Journal of Asian Finance, Economics and Business. 2022;9(5):285-94.
[12] Le TMH, Zhihong J, Zhu Z. Impact of Macroeconomic Variables on Stock Price Index: Evidence from Vietnam Stock Market. Research Journal of Finance and Accounting. 2019;10(12).
[13] Duy VQ. The Impact of Macroeconomic Factors on Stock Price Index, VN-Index. International Journal of Innovative Science, Engineering & Technology. 2016;3(7):69-84.
[14] Duy VQ, Hau LL. Impact of Macroeconomic Factors on Share Price Index in Vietnam's Stock Market. The International Journal of Engineering and Science. 2017;6(01):52-9.
[15] Mai NC. The Influence of Macroeconomic Announcements into Vietnamese Stock Market Volatility. 2016.
[16] Tien NH. The Impact of World Market on Ho Chi Minh City Stock Exchange in Context of Covid-19 Pandemic. Turkish Journal of Computer and Mathematics Education. 2021;12(14):4252-64.
[17] Nguyen TN, Nguyen DT, Nguyen VN. The Impacts of Oil Price and Exchange Rate on Vietnamese Stock Market. The Journal of Asian Finance, Economics and Business. 2020;7(8):143-50.