Forecasting Stock Valuations in Vietnam: Utilizing Statistical Models and Machine Learning Algorithm

CUONG NGUYEN CHI[1], DUY NGUYEN KHANH[2], AND TUAN ANH NGUYEN NGOC .[3]
[1]University of Information Technology, Ho Chi Minh city (email: 21520664@gm.uit.edu.vn)
[2]University of Information Technology, Ho Chi Minh city (email: 21522004@gm.uit.edu.vn)
[2]University of Information Technology, Ho Chi Minh city (email: 21521834@gm.uit.edu.vn)

ABSTRACT This research endeavors to forecast stock valuations of three prominent businesses in Vietnam, namely Petrovietnam Technical Services Corp (PVS), Asia Commercial Bank (ACB), and Vietnam Dairy Products JSC (VNM). Recognizing the pivotal role of accurate predictions in aiding investment decisions, the study employs a combination of statistical models and machine learning algorithms. The objective is to enhance the decision-making process for investors and contribute to the efficient management of financial portfolios in the Vietnamese stock market.
The research methodology involves a comprehensive examination of the historical stock prices of PVS, ACB, and VNM over the last five years within the context of the Vietnamese stock exchanges. Subsequently, leveraging statistical models and machine learning algorithms, the study seeks to provide insights into future stock valuations. This interdisciplinary approach acknowledges the transformative impact of machine learning in financial markets and aims to empower investors with robust tools for making informed and strategic investment choices.

INDEX TERMS Time Series Analysis, Machine Learning, Financial Forecasting, Vietnamese Stock Market, Random Forest, Boosting LSTM, ARIMA, ARIMAX, SVR, Linear Regression.

I. INTRODUCTION

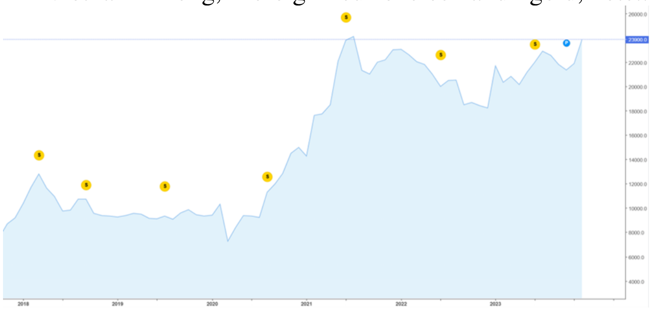
The prediction of stock price movements is a matter of great interest to investors, publicly listed companies, and governments alike. The debate on whether the market can be forecasted has persisted for a long time. According to Malkiel's (1973) Random Walk Theory, prices are arbitrarily set, making it seemingly impossible to outperform the market. However, recent breakthroughs in artificial intelligence have empirically demonstrated that stock price movements can indeed be predicted. The stock market operates as an immensely sophisticated system with vast amounts of data. Real-time data generation and constant small-scale changes are attributed to various factors and diversity within the market. Companies' ownership is dispersed among shareholders, who act as the true proprietors of the company. These shareholders own the company's shares, which are traded on a public exchange. The stocks represent ownership in the company and serve as stakeholders in its profit and loss. Given the dynamic, nonlinear, complex, nonparametric, and chaotic nature of the stock market, predicting stock market movements is considered a challenging task within the financial time series prediction process. The market is further influenced by a range of macroeconomic factors, including political developments, corporate policies, general economic conditions, investor expectations, institutional investment preferences, movements in other stock markets, and investor psychology. Concerning the Vietnamese stock markets, fluctuations parallel those of the global stock market. Recent developments in Vietnam's stock market include the revision of existing laws and regulations pertaining to control, evaluation, and transparency in financial statements and information. Forecasts predict a remarkably swift recovery for the Vietnamese stock market in 2023. Vietnam Dairy Products JSC, Asia Commercial Bank, and Petrovietnam Technical Services Corp are substantial corporations with robust presences on the Vietnamese stock exchange.

A. PETROVIETNAM TECHNICAL SERVICES CORP (PVS)

PetroVietnam is the trading name of the Vietnam Oil and Gas Group. PetroVietnam has developed rapidly since it was established in 1977, and its activities, through its various companies and wholly owned subsidiaries, now cover all the operations from oil and gas exploration and production to storage, processing, transportation, distribution and services. Wholly owned by the Vietnamese central government, it is responsible for all oil and gas resources in the country and has become its country's largest oil producer and second-largest power producer.

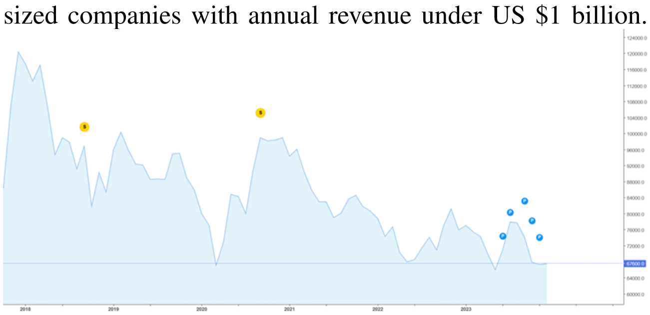
B. ASIA COMMERCIAL BANK (ACB)

Asia Commercial Bank, often abbreviated to ACB, was registered on 19 May 1993 and began operations in June 1993. Main products and services: capital mobilization (receiving deposits of customers) in Vietnam Dong, foreign currencies and gold, using capital (providing credit, investing, contributing to joint ventures) in Vietnam Dong, foreign currencies and gold, etc..



C. VIETNAM DAIRY PRODUCTS JSC (VNM)

Vinamilk, formally the Vietnam Dairy Products Joint Stock Company, is the largest dairy company in Vietnam. According to the UNDP 2007 Top 200 largest firms in Vietnam report, it was also the 15th largest company in Vietnam and formerly the most valuable public company listed in Vietnam. In 2010, it became the first company in Vietnam to be included in the Forbes Asia's 200 Best Under A Billion list, which highlights the top 200 performing small- and mid-sized companies with annual revenue under US $1 billion.



II. RELATED WORK

A stock market functions as a platform for companies to issue shares, enabling global investors to buy and sell, aiming for profit. The challenge lies in the rapid and unpredictable fluctuations of stock prices, prompting the need for predictive models. In this section, we summarize relevant literature for our research on stock price prediction, emphasizing mathematical attributes. The review covers studies proposing varied methodologies and engaging in discussions related to the common challenge of forecasting stock prices. Dias Satria's research suggests that ARIMA Box-Jenkins modeling is inadequate for stock price prediction due to the non-linear nature of the data, leading to a violation of the white noise assumption in estimating ARIMA Box-Jenkins parameters. Furthermore, in the comparison of three models—RNN, LSTM, and GRU—Satria found that the GRU model demonstrated superior performance in predicting stock prices, as evidenced by the lowest RMSE value among the three models. [10] In another study, Xiwen Jin and Chaoran Yi determined that LSTM and GRU models yielded comparatively superior results, while Random Forest showed the least favorable performance. They assessed various models based on R2 scores, revealing that LSTM achieved 0.84, GRU 0.86, Random Forest Regressor 0.51, XGBoost Regressor 0.69, Linear Regression 0.73, and LGBM Regressor 0.72. Notably, XGBoost and Random Forest exhibited subpar performance in contrast to the more effective LSTM and GRU models. [7] Furthermore, in a study conducted by Rajat Patil (2021), ARIMA, ARIMAX, and LSTM models were employed for "Time Series Analysis and Stock PriceForecasting using Machine LearningTechniques". The research findings indicate that the ARIMAX model demonstrated superior performance compared to both the ARIMA and LSTM models. [9] Bruno Miranda Henrique's study in The Journey of Finance and Data Science used Support Vector Regression (SVR) to evaluate stocks from Brazil, the U.S., and China. A linear kernel with a fixed daily training set showed reduced prediction errors in tests. It outperformed radial and polynomial kernels for daily price predictions. However, fixed training time led to decreased effectiveness as price frequency increased to minutes. In real-time prices, SVR was less accurate than a random walk model under fixed training. [4] The Computer Science Department at Bina Nusantara University in Jakarta, Indonesia, developed a stock price prediction program using LSTM implemented with Python and Tensorflow. The program achieved an impressive accuracy of 94.59% after 100 epochs. In comparison to other research on stock price forecasting, the LSTM method exhibited superior performance, surpassing typical neural network methods that usually achieve around 90% accuracy. [2]

III. MATERIALS & METHODS

A. DATA COLLECTION

We obtained a dataset containing information on the stock market performance of three prominent Vietnamese companies—Petrovietnam Technical Services Corp (PVS), Asia Commercial Bank (ACB), and Vietnam Dairy Products JSC (VNM)—from the investing.com website. Investing.com is a comprehensive platform offering stock trading information for companies globally, including those in Vietnam. Each dataset includes the following information:

- **Date**: The opening day of stock trading.
- **Price (Close Price)**: The last price at which a stock trades at the end of the exchange.
- **Open**: The initial price at which a stock opens for trading.
- **High**: The highest stock price recorded during the day.
- **Low**: The lowest stock price recorded during the day.
- **Volume**: The number of shares traded, representing the buying and selling activities.
- **Change**: The percentage change in the stock price from the previous day, indicating today's performance.

B. DESCRIPTIVE STATISTICS

	ACB	PVS	VNM
Count	1511	1516	1514
Max	30360.0	40700.0	175578.0
Min	8763.1	9000.0	61260.1
Mean	18031.30	22345.78	94329.12
Median	17504.0	21900.0	88061.1
Std	6257.79	6612.24	25267.59
Coefficient of Deviation	0.35	0.30	0.27
Variance	39159934.62	43721677.18	638451010.37
Skewness	0.14	0.56	1.53
Kurtosis	-1.62	-0.00	1.90
25%	11626.0	17929.0	76361.4
50%	17504.0	21900.0	88061.1
75%	24150.0	25738.0	101599.8

TABLE 1. Descriptive Statistics

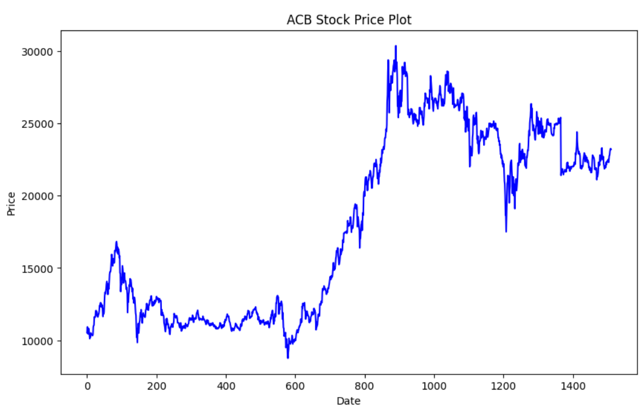


FIGURE 1. ACB stock price plot

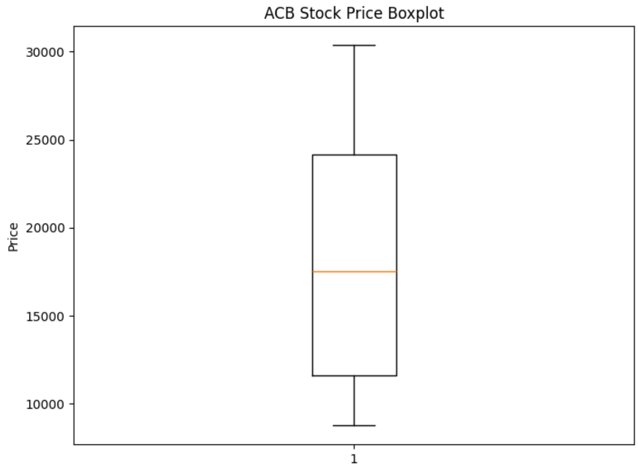


FIGURE 2. ACB stock price box plot

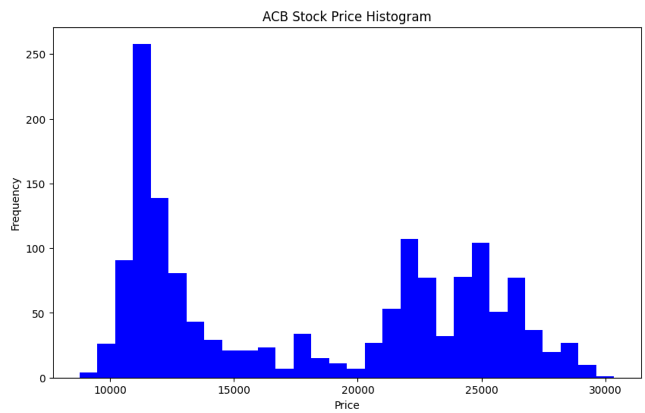


FIGURE 3. ACB stock price histogram plot

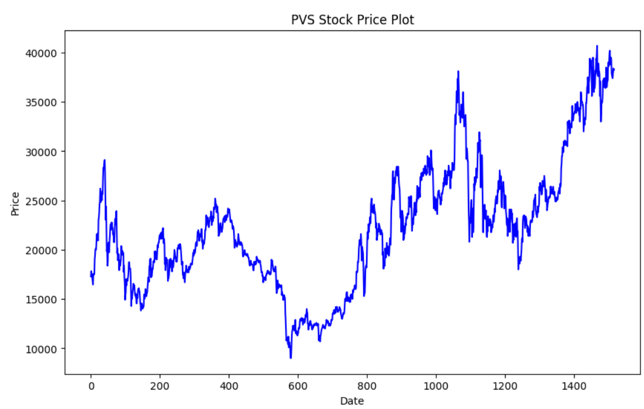


FIGURE 4. PVS stock price plot

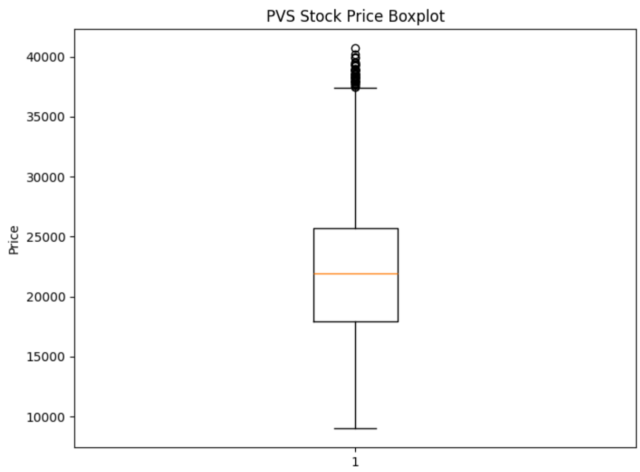


FIGURE 5. PVS stock price box plot

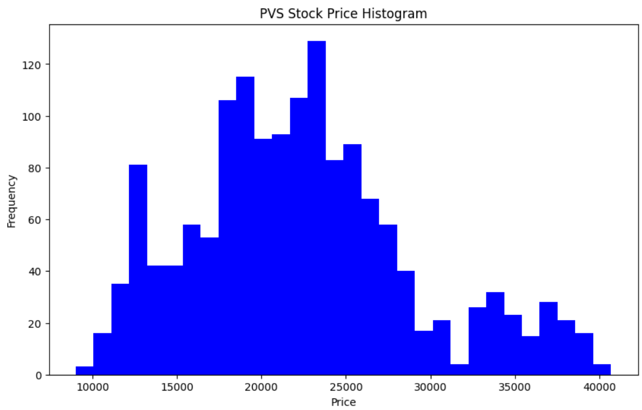


FIGURE 6. PVS stock price histogram plot

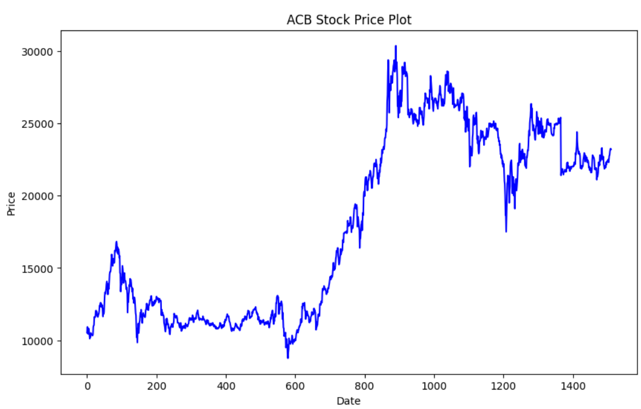


FIGURE 7. ACB stock price plot

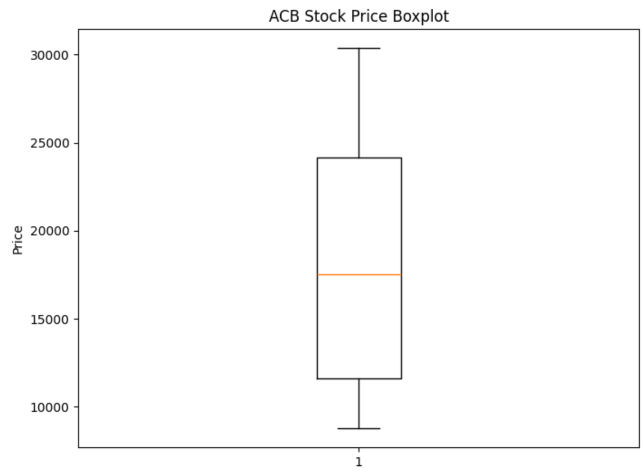


FIGURE 8. ACB stock price box plot

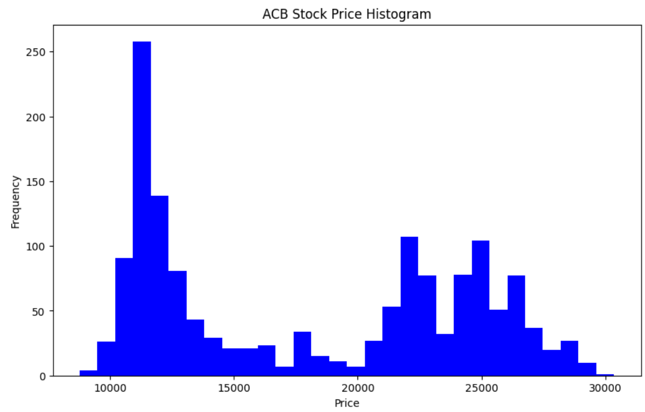


FIGURE 9. ACB stock price histogram plot

C. ALGORITHMS

1) Linear Regression

Regression analysis is a tool for building mathematical and statistical models that characterize relationships between a dependent variable and one or more independent, or explanatory, variables, all of which are numerical. This statistical technique is used to find an equation that best predicts the y variable as a linear function of the x variables. A multiple linear regression model has the form: [**?**]

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

Where:

- Y is the dependent variable.
- $X_1, \ldots, X_k$ are the independent (explanatory) variables.
- β_0 is the intercept term.
- $\beta_1, \ldots, \beta_k$ are the regression coefficients for the independent variables.
- ε is the error term.

2) ARIMA

An ARIMA model is a class of statistical models for analyzing and forecasting time series data.

The parameters of the ARIMA model are:

- p: The number of lags
- d: The number of times that the raw observations are differenced.
- q: The size of the moving average window.

The ARIMA model equation is given by:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$

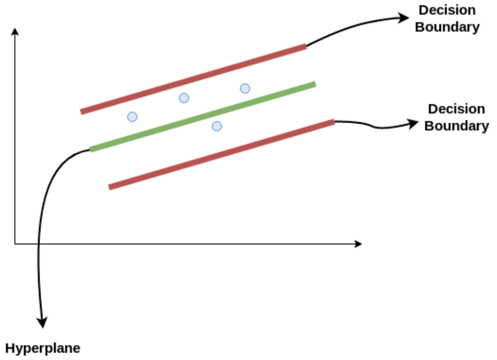
3) Support Vector Regression (SVR)

Support Vector Regression (SVR) is a type of machine learning algorithm used for regression analysis. The goal of SVR is to find a function that approximates the relationship between the input variables and a continuous target variable, while minimizing the prediction error.

Unlike Support Vector Machines (SVMs) used for classification tasks, SVR seeks to find a hyperplane that best fits the data points in a continuous space. This is achieved by mapping the input variables to a high-dimensional feature space and finding the hyperplane that maximizes the margin (distance) between the hyperplane and the closest data points, while also minimizing the prediction error.

SVR can handle non-linear relationships between the input variables and the target variable by using a kernel function to map the data to a higher-dimensional space. This makes it a powerful tool for regression tasks where there may be complex relationships between the input variables and the target variable.

Support Vector Regression (SVR) uses the same principle as SVM, but for regression problems.



Consider these two red lines as the decision boundary, and the green line as the hyperplane. Our objective, when we are moving on with SVR, is to basically consider the points that are within the decision boundary line. Our best fit line is the hyperplane that has a maximum number of points.

The first thing that we'll understand is what is the decision boundary (the danger red line above!). Consider these lines as being at any distance, say $'a'$, from the hyperplane. So, these are the lines that we draw at distance $'+a'$ and $'-a'$ from the hyperplane. This $'a'$ in the text is basically referred to as epsilon.

Assuming that the equation of the hyperplane is as follows:

$$Y = wx + b$$

(equation of hyperplane) Then the equations of the decision boundary become:

$$wx + b = +a$$

$$wx + b = -a$$

Thus, any hyperplane that satisfies our SVR should satisfy:

$$-a < Y - wx + b < +a$$

Our main aim here is to decide a decision boundary at $'a'$ distance from the original hyperplane such that data points closest to the hyperplane or the support vectors are within that boundary line.

Hence, we are going to take only those points that are within the decision boundary and have the least error rate or are within the Margin of Tolerance. This gives us a better fitting model.

4) The Long Short-Term Memory (LSTM) Model

The Long Short-Term Memory (LSTM) model is a type of recurrent neural network (RNN) specifically designed to handle long-term dependencies in sequential data. It was introduced by Hochreiter and Schmidhuber in 1997 and has since gained popularity in various domains, including natural language processing, speech recognition, and time series forecasting.

Unlike standard feedforward neural networks, LSTM incorporates feedback connections, making it a recurrent neural network (RNN). This enables LSTM to not only process individual data points like images but also handle sequential data such as speech or video. LSTM, being a special type of RNN, has demonstrated exceptional performance across a wide range of problem domains. Moreover, LSTM is designed to address the issue of long-term dependencies. It has the inherent ability to remember information over extended periods without the need for explicit training. This means that LSTMs can retain information by default, without any additional intervention.

Every recurrent neural network (RNN) has a sequential structure consisting of repeating modules of neural networks. In standard RNNs, these modules have a very simple structure, often a single layer.

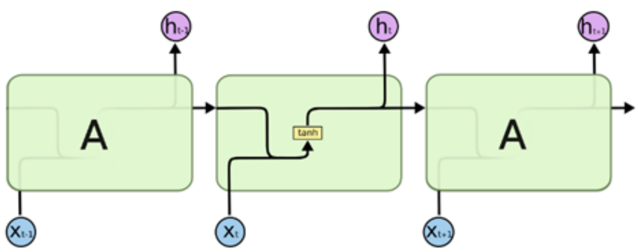


FIGURE 10. The repeating module in a standard RNN contains a single layer

LSTM also has a similar sequential structure, but its modules have a unique structure with 4 interacting layers.

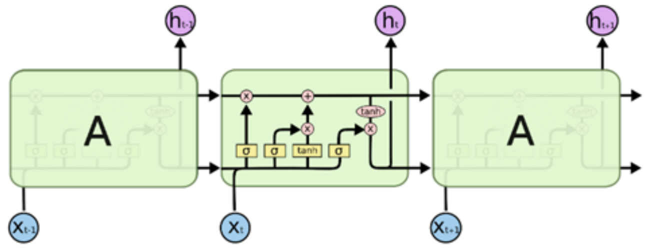


FIGURE 11. The LSTM model architecture

5) Random Forests

Random forests represent a machine learning model that amalgamates multiple decision trees to yield more precise and resilient outcomes compared to a single tree. This approach leverages the "wisdom of the crowds," positing that the collective insights of many simple models can surpass the accuracy of a more intricate model [**?**]. The functioning of random forests unfolds in four steps:

1) Randomly selecting samples from the given dataset.
2) Constructing a decision tree for each sample and extracting prediction results from each tree.
3) Collecting votes for each prediction result.
4) Identifying the most frequently predicted result as the final prediction.

Decision trees, fundamental machine learning models applicable to classification or forecasting tasks, operate by partitioning data into branches based on specific features. Random forests employ multiple decision trees to enhance result accuracy.

The process involves the following steps:

1) For each iteration from $k = 1$ to K:
   - Sampling a bootstrap sample from the training data.
   - Training a decision tree T_k on the bootstrap sample.
2) For a new input X:
   - For each tree T_k, obtaining the prediction Y_k.
   - Aggregating predictions: $\hat{Y} = \frac{1}{K}\sum_{k=1}^{K} Y_k$ (for regression) or $\hat{Y} = \text{mode}(Y_1, Y_2, \ldots, Y_K)$ (for classification).

6) XGBoost-LSTM

XGBoost, short for eXtreme Gradient Boosted trees, is a powerful machine learning algorithm employing an ensemble of decision trees and gradient boosting for predictions. Widely utilized in data science, it has demonstrated success in numerous machine learning competitions. [8]

XGBoost efficiently addresses classification and regression problems, with applicability extended to time series forecasting by transforming the dataset into a supervised learning problem. The algorithm operates as an optimized, regularized gradient-boosted tree, renowned for its high performance and suitability for big data processing. [3]

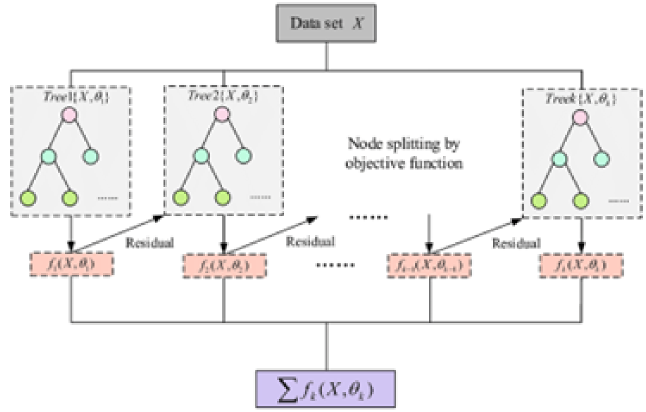


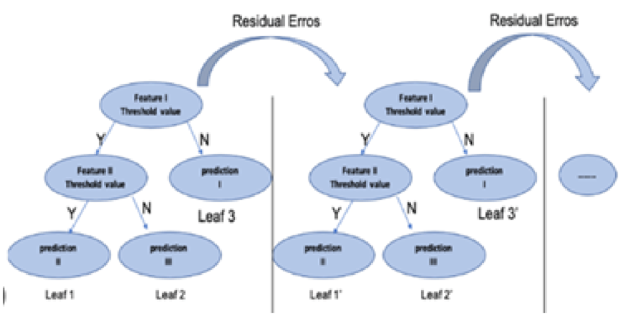
FIGURE 12. Model of XGBoost Architectural

Boosting, a fundamental concept behind XGBoost, involves sequentially training base models, with each subsequent model focusing on challenging examples misclassified by prior models. XGBoost builds a sequence of decision trees, combining their outputs to generate predictions. The final prediction is computed by summing up the tree outputs and applying an activation function if necessary.

The regularized learning objective of XGBoost involves a regularization term, contributing to its optimization process. The objective function combines the training loss and regularization. [11]

To enhance predictive models, XGBoost and LSTM [5] [1] can be combined using the "Ensemble Model" approach:

1) Train an LSTM model and an XGBoost model independently.
2) Combine predictions from both models using averaging or weighting.
3) Utilize the combined model to make predictions on test data.



7) Fully Convolutional Neural Networks (FCN)

Fully Convolutional Neural Networks (FCN) were first proposed in Wang et al. (2017b) for classifying univariate time series and validated on 44 datasets from the UCR/UEA archive. FCNs are mainly convolutional networks that do not contain any local pooling layers, which means that the length of a time series is kept unchanged throughout the convolutions. In addition, one of the main characteristics of this architecture is the replacement of the traditional final fully connected (FC) layer with a Global Average Pooling (GAP) layer. This GAP layer drastically reduces the number of parameters in a neural network while enabling the use of the Class Activation Map (CAM) (Zhou et al., 2016). The CAM highlights which parts of the input time series contributed the most to a certain classification. [6]

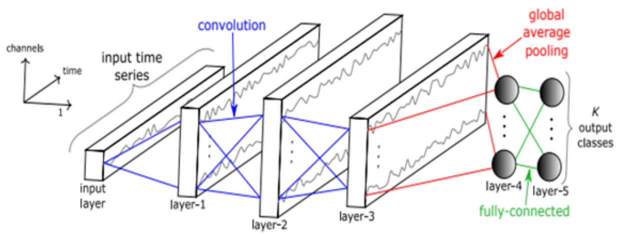


FIGURE 13. FCN Architecture

IV. EVALUATION

To assess the models' accuracy in my case, three metrics, namely Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Directional Accuracy (MDA), are employed. The algorithm exhibiting the lowest values for these metrics demonstrates superior performance. The formulas for RMSE, MAPE, and MDA are as follows:

1) **Root Mean Squared Error (RMSE):**

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

2) **Mean Absolute Percentage Error (MAPE):**

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$

3) **Mean Directional Accuracy (MDA):**

$$MDA = \frac{1}{N}\sum_{t=1}^{N}\text{sign}(X_t - X_{t-1}) == \text{sign}(F_t - X_{t-1})$$

where:

MDA is the Mean Directional Accuracy,

N is the total number of observations or time periods,

t is a time index ranging from 1 to N,

X_t denotes the value of a variable at time t,

F_t represents the forecasted value at time t,

$\text{sign}(\cdot)$ is the sign function,

Mean Directional Accuracy (MDA) is a metric used to assess the accuracy of directional predictions made by a forecasting model. It is particularly relevant in situations where the model's primary task is to predict the direction of a future outcome, such as whether a financial market will go up or down. The MDA metric provides a percentage that represents the average accuracy of the model in correctly predicting the direction of the outcome. A higher MDA value indicates better performance in terms of directional accuracy.

V. RESULT

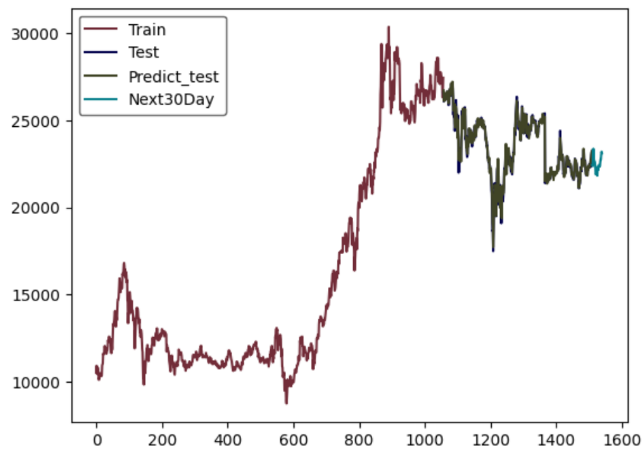
In this section, we use the best performance (lowest error) model for each train/test ratio of each dataset to predict the next 30 days.

A. ACB DATASET

1) 7:3 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	33.98	8389.86	46.90
SVR	1.28	458.68	0.42
LSTM	1.79	597.73	42.42
Linear Regression	18.73	4957.35	46.90
ARIMAX	0.60	188.31	69.09
Random Forest	3.69	1048.88	41.94
XGBoost-LSTM	2148410.94	499592420.0	42.89
FCN	7.03	1814.70	43.50

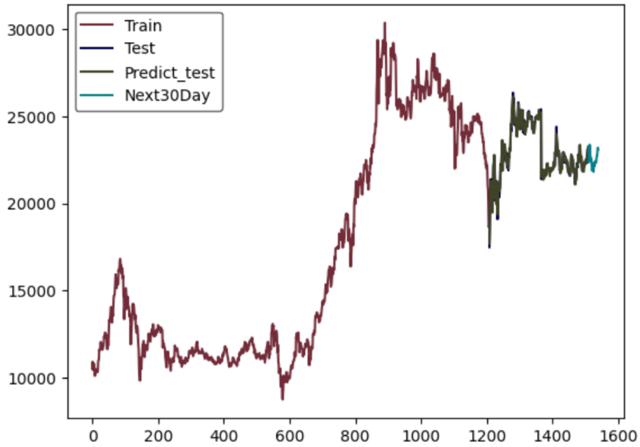
TABLE 2. Performance Metrics for 7:3 Ratio



2) 8:2 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	18.11	4467.50	41.53
SVR	1.25	453.12	0.43
LSTM	1.30	448.33	40.96
Linear Regression	23.27	5595.79	47.51
ARIMAX	0.57	178.86	68.54
Random Forest	1.25	440.77	44.28
XGBoost-LSTM	2146750.0	497644100.0	44.28
FCN	2.67	785.11	39.71

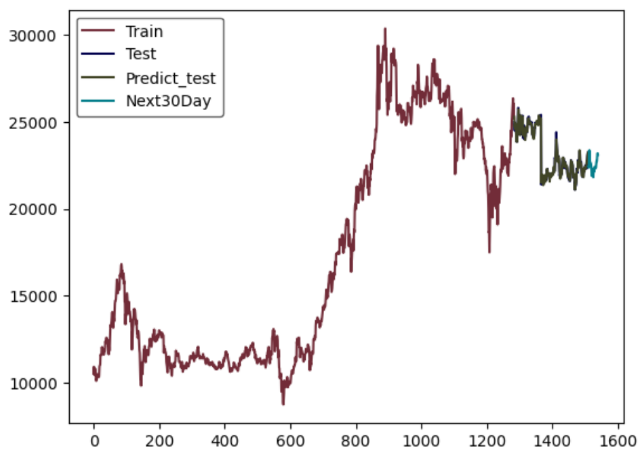
TABLE 3. Performance Metrics for 8:2 Ratio



3) 8.5:1.5 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	9.32	2438.23	49.78
SVR	0.94	388.30	0.40
LSTM	0.91	386.33	40.00
Linear Regression	19.76	4901.18	44.00
ARIMAX	0.43	127.80	64.16
Random Forest	3.69	1048.88	41.94
XGBoost-LSTM	2171251.95	498092960.0	38.46
FCN	61.66	14205.71	13.78

TABLE 4. Performance Metrics for 8.5:1.5 Ratio

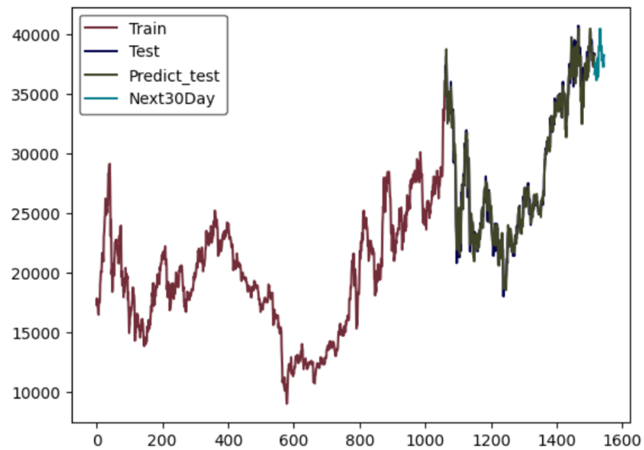


B. PVS DATASET

1) 7:3 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	32.53	9432.26	7.51
SVR	11.53	6450.24	0.43
LSTM	3.49	1223.76	43.03
Linear Regression	19.05	8084.05	48.57
ARIMAX	1.32	483.27	67.40
Random Forest	3.99	1750.38	32.39
XGBoost-LSTM	3121189.65	900733600.0	42.79
FCN	4.50	1624.99	44.10

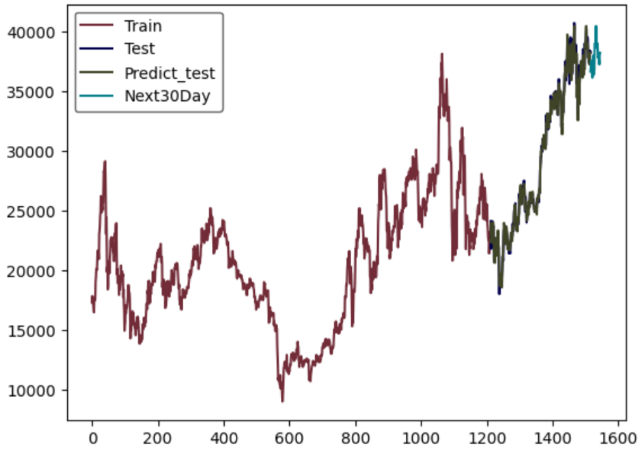
TABLE 5. Performance Metrics for 7:3 Ratio



2) 8:2 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	21.77	9258.59	8.94
SVR	5.29	3190.50	0.39
LSTM	2.02	807.43	41.91
Linear Regression	16.92	7065.51	49.01
ARIMAX	1.13	412.92	63.04
Random Forest	3.62	1771.39	48.53
XGBoost-LSTM	3205419.14	993224500.0	42.65
FCN	3.45	1347.05	42.12

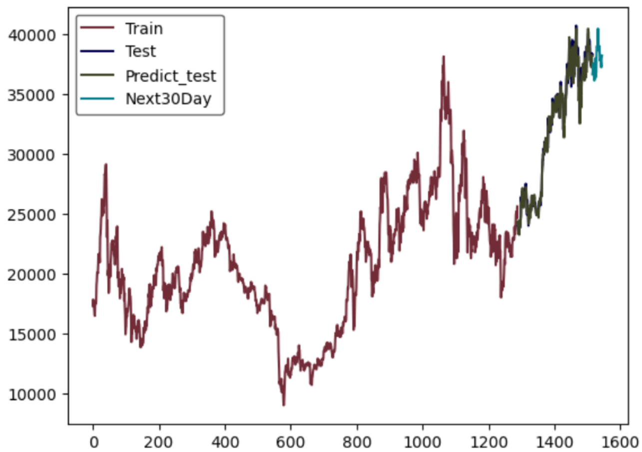
TABLE 6. Performance Metrics for 8:2 Ratio



3) 8.5:1.5 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	19.17	8311.35	8.85
SVR	6.24	3717.92	0.39
LSTM	2.86	1206.37	40.82
Linear Regression	20.19	8522.40	50.44
ARIMAX	1.00	418.05	61.23
Random Forest	3.99	1750.38	32.39
XGBoost-LSTM	3195078.91	1068951600.0	39.80
FCN	4.99	2027.48	46.19

TABLE 7. Performance Metrics for 8.5:1.5 Ratio

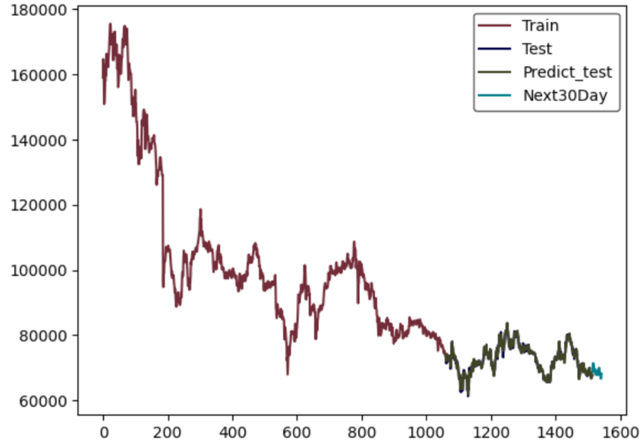


C. VNM DATASET

1) 7:3 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	5.19	4539.97	47.24
SVR	10.57	9057.88	0.45
LSTM	1.92	1734.17	46.57
Linear Regression	24.52	20355.51	50.33
ARIMAX	0.53	504.91	77.53
Random Forest	5.02	4542.53	31.44
XGBoost-LSTM	11464873.44	8346236400.0	47.28
FCN	15.45	12242.99	6.37

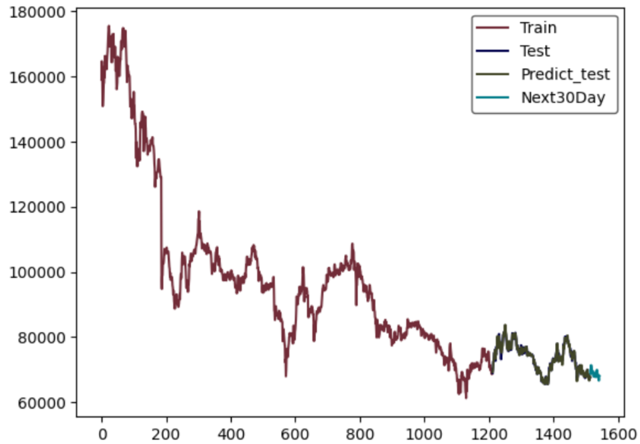
TABLE 8. Performance Metrics for 7:3 Ratio



2) 8:2 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	6.75	6164.57	47.18
SVR	1.75	1654.43	0.42
LSTM	1.40	1311.68	45.02
Linear Regression	27.60	20940.81	48.50
ARIMAX	0.51	502.40	75.50
Random Forest	1.23	1138.23	43.17
XGBoost-LSTM	11578707.03	8527116000.0	45.02
FCN	16.40	13000.77	6.25

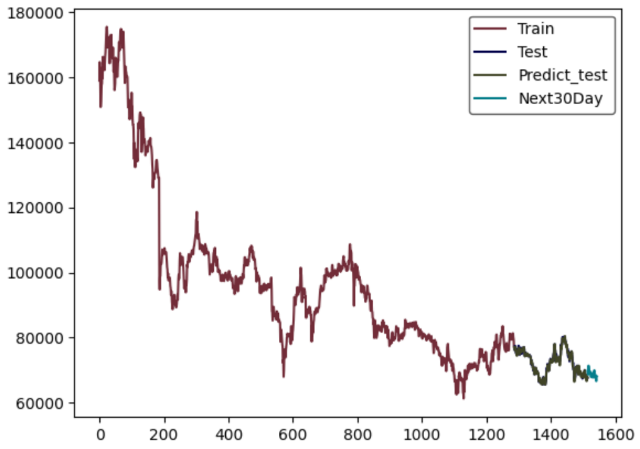
TABLE 9. Performance Metrics for 8:2 Ratio



3) 8.5:1.5 Train/Test ratio

Model	MAPE	RMSE	MDA
ARIMA	7.32	6135.52	24.78
SVR	1.89	1772.10	0.42
LSTM	1.41	1290.04	45.92
Linear Regression	23.24	17534.68	51.77
ARIMAX	0.45	422.04	74.89
Random Forest	5.02	4542.53	31.44
XGBoost-LSTM	11426432.81	8224911000.0	44.90
FCN	1.95	1757.89	50.25

TABLE 10. Performance Metrics for 8.5:1.5 Ratio



VI. CONCLUSION

The application of ARIMAX models in forecasting stock prices for major Vietnamese banks like ACB, PVS, and VNM has shown significant promise in learning and generating forecasts. These models, alongside others employed in this project such as ARIMA, SVR, LSTM, Linear Regression, Random Forest, XGBoost-LSTM, and FCN, demonstrate the ability to handle time series data and capture intricate relationships within sequences. When assessing the predictions of stock prices for the next 30 days, ARIMAX models suggest a relatively stable trend in the future stock prices of these banks. However, it is essential to emphasize that determining the "suitability" of a model to predict an upward trend requires not only evaluating its predictive capability but also a deep understanding of the financial market. The fine-tuning and optimization of the model play a crucial role in achieving accurate predictions. Therefore, it is crucial to recognize that prediction accuracy is not solely derived from the model itself. The effectiveness of predictions also depends on the seamless integration of the model with domain expertise and the process of adjusting the model for practical real-world applications.

VII. ORIENTATION

Our current models, while valuable, may not be entirely accurate, necessitating the allocation of time for validating real-world predictions. Potential errors within these models could stem from research limitations or inappropriate model selection concerning the utilized datasets. Therefore, there exists an opportunity for ongoing improvement in our models for future applications. One avenue for enhancing our existing models involves the integration of attention mechanisms into ARIMAX models. This addition allows a focused analysis of critical parts of the input sequence, thereby improving the transformation from input to output data. Additionally, our strategy includes exploring alternative models such as Vector Autoregression (VAR), Sequence to Sequence (Seq2Seq), and various Gradient Boosting methods. The objective is to compare their performance with our current models in practical scenarios, conducting a comprehensive analysis to identify the most suitable model for our specific purposes.

In conclusion, our approach combines the enhancement of existing models with the exploration of new ones, aiming to improve accuracy and efficacy in real-world applications.

ACKNOWLEDGMENT

We wish to convey our utmost appreciation to Assoc. Prof. Dr. Nguyen Dinh Thuan for his invaluable guidance, steadfast support, and insightful mentorship throughout the development of this paper. His profound expertise and unwavering encouragement have significantly contributed to the success of this research endeavor. Our sincere gratitude extends to TA. Nguyen Minh Nhut for his dedicated assistance, thoughtful insights, and meticulous attention to detail during the various stages of this project. His unwavering commitment and collaborative spirit have played a pivotal role in shaping the overall quality of the work. We consider ourselves truly fortunate to have had the privilege of collaborating with Assoc. Prof. Dr. Nguyen Dinh Thuan and TA. Nguyen Minh Nhut. We express deep gratitude for their substantial contributions, which have not only enriched the depth but also bolstered the credibility of this paper.

REFERENCES

[1] Shivanshu Aggarwal. The ultimate guide to building your own lstm models. https://www.projectpro.io/article/lstm-model/832, November 2023.
[2] Widodo Budiharto. Data science approach to stock prices forecasting in indonesia during covid-19 using long short-term memory (lstm). Journal of Big Data, 8:1–9, 2021.
[3] Distributed (Deep) Machine Learning Community. Introduction to boosted trees. https://xgboost.readthedocs.io/en/stable/tutorials/model.html, December 2023.
[4] Bruno Miranda Henrique, Vinicius Amorim Sobreiro, and Herbert Kimura. Stock price prediction using support vector regression on daily and up to the minute prices. The Journal of Finance and Data Science, 4(3):183–201, 2018.
[5] Sepp Hochreiter. Long short-term memory. https://www.researchgate.net/publication/13853244_Long_Short-term_Memory, December 1997.
[6] H. Ismail Fawaz, G. Forestier, and J. et al. Weber. Deep learning for time series classification: a review. Data Min Knowl Disc, 02 March 2019.
[7] X. Jin and C. Yi. The comparison of stock price prediction based on linear regression model and machine learning scenarios. In 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022), page 837–842. Atlantis Press, December 2022.
[8] NVIDIA. Xgboost. https://www.nvidia.com/en-us/glossary/data-science/xgboost/, December 2023. Accessed Dec. 22, 2023.
[9] R. Patil. Time series analysis and stock price forecasting using machine learning techniques. Technical report, February 2021.
[10] D. Satria. Predicting banking stock prices using rnn, lstm, and gru approach. Appl. Comput. Sci., 19:82–94, March 2023.
[11] Great Learning Team. Understanding xgboost algorithm | what is xgboost algorithm? https://www.mygreatlearning.com/blog/xgboost-algorithm/, June 2022.