FORECASTING STOCK MARKET PRICES USING MACHINE LEARNING AND DEEP LEARNING MODELS

NGO VAN MANH[1], NGUYEN NGOC THANH[2], AND TRAN QUOC KHANG[3]

[1]Faculty of Information Systems, University of Information Technology, (e-mail: 21522328@gm.uit.edu.vn)
[2]Faculty of Information Systems, University of Information Technology, (e-mail: 21522600@gm.uit.edu.vn)
[3]Faculty of Information Systems, University of Information Technology, (e-mail: 21522200@gm.uit.edu.vn)

ABSTRACT This paper conducts an analysis and stock price forecasts for VCB, BID, and CTG, the top three banks in Vietnam, crucial entities in the financial sector. These banks play a pivotal role in the country's economic landscape. Leveraging a diverse set of models including Linear Regression, ARIMA, RNN, GRU, LSTM, and N-HITS, we predict both the highest and lowest stock prices of these banks. Furthermore, we employ optimization algorithms such as Nadam and Adadelta to enhance the performance of our Deep learning model and compare when not using optimization models to give insights on efficiency. By scrutinizing stock price fluctuations and employing advanced forecasting methodologies, this study provides valuable insights for investors and stakeholders in the Vietnamese banking industry.

INDEX TERMS Stock price forecasting, Vietnamese banks, VCB, BID, CTG, Linear Regression, ARIMA, RNN, GRU, LSTM, N-HITS, Nadam, Adadelta.

I. INTRODUCTION

The stock market refers to the collection of markets and ex-change centers where economic activities like buying, selling,and deploying shares of publicly held companies take place.Such financial practices are conducted through institutionalized formal exchanges through over-the-counter marketplaces that function under a defined set of regulations. The stock market is a very dynamic and uncertain field, so the stock market prediction naturally becomes a burning topic. Due to the advancement of computational power in recent times, pre-dicting the stock market has been much faster and accurate.Artificial Intelligence and machine learning models play a crucial role in predicting stock prices and, hence, determining an accurate result [1]

This study chooses the Ho Chi Minh Stock Exchange (HOSE) in VietNam, including three different stock market of three VietNamese Banks. There are many algorithms and techniques that help us predict prices. In this paper, we will use Linear Regression, LSTM, RNN, GRU, N-HITS models to predict the stock prices of VietcomBank, BIDV, VietinBank for the next 30 days

II. RELATED WORKS

This section reviews relevant studies that explore the application of mathematical attributes for stock price prediction. In this study, we will use the following 8 algorithms: Linear regression, ARIMA, RNN, GRU, LSTM, N-HiTS, Nadam, Adadelta to forecast stock prices in Vietnam. In which Nadam & Adadelta are two optimal algorithms.

A Multi Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model, research by Shahzad Zaheer et al. In this article, they use LTSM, Deep Learning Model to forecast stock prices. But what's special here is that the model takes the original stock data and gives two parameters: closing price and high price for the next day. This proposed method outperforms existing stock price forecasting methods and is considered to be the most accurate and suitable for short-term forecasting. [2]

A Comparative Research of Stock Price Prediction of Selected Stock Indexes and the Stock Market by Using Arima Model, research by Nayab Minhaj, Roohi Ahmed, Irum Abdul Khalique, Mohammad Imran. This paper explores the use of the ARIMA model for short-term prediction of Johnson & Johnson (JNJ) stock prices, demonstrating its effectiveness compared to traditional methods. [3]

NHITS: Neural Hierarchical Interpolation for Time Series Forecasting, research by Cristian Challu et al. In this paper, the author tested large-scale time series datasets for long-

term forecasting, demonstrating the advantages of NHITS providing an average accuracy improvement of nearly 20% compared to the latest Transformer At the same time, it also reduces calculation time. [4]

Nadam: A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer, Research by Jatin Sharma et al long-term solar photovoltaic power forecasting using LSTM with Nadam optimizer. The proposed model is compared with two time series models and eight artificial neural network models using LSTM with different optimizers. The obtained results using LSTM with Nadam optimizer present a significant improvement in the forecasting accuracy of 30.56% over autoregressive integrated moving average, 47.48% over seasonal autoregressive integrated moving average, and 1.35%, 1.43%, 3.51%, 4.88%, 11.84%, 50.69%, and 58.29% over models using RMSprop, Adam, Adamax, SGD, Adagrad, Adadelta, and Ftrl optimizer, respectively. [5]

Adelta: ADADELTA: AN ADAPTIVE LEARNING RATE METHOD, Research by Matthew D. Zeiler. Adadelta is a per-dimension learning rate method for gradient descent, dynamically adjusting rates during training with minimal computation. It eliminates manual tuning and proves robust to noise, different architectures, data types, and hyperparameters. The method shows promising results on MNIST and large-scale voice datasets, effective in both single-machine and distributed settings. [8]

III. MATERIALS AND METHODOLOGY

A. DATASET

The historical stock price of Joint Stock Commercial Bank for Foreign Trade of Vietnam (VCB), Bank for Investment and Development of Vietnam (BIDV) and Military Commercial Joint Stock Bank (CTG) from 01/01/2015 to 27/03/2024 will be applied. The data contains column such as Date, Price, Open, High, Low, Close, Adj Close, Volume. As the goal is to forecast high and low prices, data relating to column “High", "Low" (VND) will be processed.

B. DESCRIPTIVE STATISTICS

TABLE 1. VCB, BIDV, CTG's Descriptive Statistics

High	VCB	BID	CTG
Count	2295	2301	2301
Mean	49,307	25,702	19,535
Std	22,022	10,249	6,964
Min	15,779	9,172	9,707
25%	25,56	15,311	13,625
50%	49,305	26,705	16,571
75%	66,293	32,385	25,819
Max	100,5	56,7	38,066

Low	VCB	BID	CTG
Count	2295	2301	2301
Mean	48,157	24,956	18,997
Std	21,587	9,978	6,767
Min	15,435	9,031	9,568
25%	25,069	14,817	13,243
50%	47,778	25,753	16,051
75%	64,804	31,256	25,192
Max	97,3	53,5	37,303

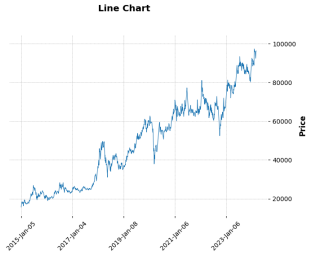

FIGURE 1. Vietcombank stock price's line chart

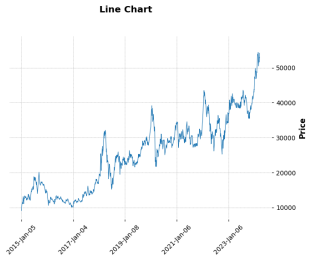

FIGURE 2. BIDV stock price's line chart

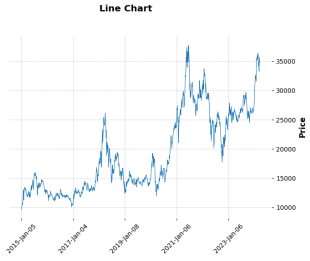

FIGURE 3. Vietinbank stock price's line chart

C. METHODOLOGY

IV. METHODOLOGY

A. LINEAR REGRESSION

Regression models are used for describing relationships between variables by fitting a line to the observed data. Regression can estimate how a dependent variable changes as the independent variables change. Multiple linear regression is used for estimating the relationship between two or more independent variables and one dependent variable A multiple linear regression model has the form: [7]

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

Where:

- Y is the dependent variable (Target Variable).
- $X_1, X_2, \ldots, X_k$ are the independent (explanatory) variables.
- β_0 is the intercept term.
- $\beta_1, ..., \beta_k$ are the regression coefficients for the independent variables.
- ε is the error term.

B. ARIMA

An autoregressive integrated moving average (ARIMA) model is a statistical tool utilized for analyzing time series data, aimed at gaining deeper insights into the dataset or forecasting forthcoming trends. [**?**]

Autoregressive (AR)

AR is Auto Regression, and p is the number of autoregressive terms . The equation for AR model is:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \delta + \varepsilon_t$$

Moving Average (MA)

MA is the Moving Average, and q is the number of terms in the moving average. The equation for MA model is:

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_p \varepsilon_{t-p} + \mu + \varepsilon_t$$

Differencing (I)

Last, the I part is Integrated, and d is the number of differences (order) required to make it a stationary sequence. For example:

$$If d = 0 : \Delta Y_t = Y_t$$

$$If d = 1 : \Delta Y_t = Y_t - Y_{t-1}$$

$$If d = 2 : \Delta Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$$

After combining them, we will have the ARIMA (p, d, q) express as follow:

$$\Delta Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_p \varepsilon_{t-p} + \varepsilon_t$$

REFERENCES

[1] Somenath Mukherjee, Bikash Sadhukhan, Stock market prediction using deep learning algorithms, ResearchGate, https://www.researchgate.net/publication/354268488_Stock_market_prediction_using_deep_learning_algorithms

[2] Shahzad Zaheer, Nadeem Anjum, Saddam Hussain, Abeer D. Algarni, Jawaid Iqbal, Sami Bourouis, Syed Sajid Ullah, A Multi Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model, https://www.mdpi.com/2227-7390/11/3/590

[3] Nayab Minhaj, Roohi Ahmed, Irum Abdul Khalique, Mohammad Imran, A Comparative Research of Stock Price Prediction of Selected Stock Indexes and the Stock Market by Using Arima Model, https://ojs.wiserpub.com/index.php/GES/article/view/1426/1046

[4] Cristian Challu, Kin G. Olivares, Boris N. Oreshkin, Federico Garza Ramirez, Max Mergenthaler Canseco, Artur Dubrawski, NHITS: Neural Hierarchical Interpolation for Time Series Forecasting, https://ojs.aaai.org/index.php/AAAI/article/view/25854

[5] Jatin Sharma1, Sameer Soni, Priyanka Paliwal, Shaik Saboor, Prem K. Chaurasiya, Mohsen Sharifpur, Nima Khalilpoor, Asif Afzal, A novel long term solar photovoltaic power forecastingapproach using LSTM with Nadam optimizer: A case studyof India, https://onlinelibrary.wiley.com/doi/epdf/10.1002/ese3.1178

[6] Matthew D. Zeiler, ADADELTA: AN ADAPTIVE LEARNING RATE METHOD, https://browse.arxiv.org/pdf/1212.5701v1

[7] Rebecca Bevans, Multiple Linear Regression, Published on February 20,2020,https://www.scribbr.com/statistics/multiple-linear-regression/

[8] Duke University, ARIMA models for time series forecasting, https://people.duke.edu/~rnau/411arim.htm