FFT, TimesNet, and Random Forest in Real Estate Stock Market Analysis

PHAN CHI CUONG[1], NGUYEN LE KHANG[2], AND LE PHAM QUOC BAO[3]

[1]Faculty of Information Systems, University of Information Technology, (e-mail: 21520673@gm.uit.edu.vn)
[2]Faculty of Information Systems, University of Information Technology, (e-mail: 21520960@gm.uit.edu.vn)
[3]Faculty of Information Systems, University of Information Technology, (e-mail: 21521849@gm.uit.edu.vn)

ABSTRACT This study investigates the effectiveness of Fast Fourier Transform (FFT), Time Series Network (TimeSNet), and Random Forest (RF) models in predicting stock prices within the Vietnamese real estate market. We apply these models independently to historical daily closing prices of three major real estate companies from 2019 to 2024, exploring how each method contributes to understanding and forecasting stock price movements. FFT is utilized to reveal underlying periodic patterns, TimeSNet to capture temporal dependencies, and RF to provide robust predictions. The results offer insights into the strengths and weaknesses of each model for this specific market, providing valuable information for investors and policymakers.

INDEX TERMS Placeholder

I. INTRODUCTION

Time-series forecasting plays a crucial role in decision-making across various domains. Its significance lies in its ability to provide valuable insights into future trends and patterns in time-dependent data. For instance, accurate predictions of stock prices, interest rates, and foreign exchange rates are essential for informed investment decisions in finance. Similarly, healthcare organizations rely on forecasting patient demand and resource utilization to allocate resources effectively and improve patient care. Energy management companies use time series forecasting to optimize energy production, distribution, and consumption. The accuracy and efficiency of time-series forecasting models significantly impact organizational performance and decision-making processes.

In this paper, we explore an innovative approach to enhance time-series forecasting using the Fast Fourier Transform (FFT). The FFT algorithm extracts frequency-domain features from time series data, offering a promising avenue for improving forecast accuracy and computational efficiency. Our investigation involves a comparative analysis of models trained with FFT-based features against traditional time domain features. We apply this approach to predict stock prices of real estate companies, leveraging not only FFT but also other techniques such as TimesNet and Random Forest. Through our study, we shed light on the interpretability of frequency domain features and their relationship with underlying time series patterns, emphasizing the potential of FFT-based feature engineering in enhancing forecasting models.

II. RELATED WORKS

In recent years, many stock prediction models have been researched and many articles have been published, such as:

Hind Daori, Alanoud Alanazi, Manar Alharthi, Ghaida Alzahrani (2022) [1] used Artificial Neural Network (ANN), Random Forest Classifier, Logistic Regression,and then analyze and predict the patterns of previous stock prices and the results showed that the models were efficient and produced better results.

Hugo Souto(2023) [2] has researched about TimesNet for Realized Volatility Prediction. Finally, they concluded that TimesNet stands out as a reliable and effective benchmark model for researching realized volatility. Although it may not always surpass NBEATSx and NHITS in every metric, its strong performance and consistency make it a valuable option, especially when compared to TFT. Overall, TimesNet presents a balanced and dependable choice that combines reliability with effectiveness, making it a suitable neural network model for researchers and practitioners in the field of realized volatility.

In another article by Bohumil Stádník, Jurgita Raudeliuniene, Vida Davidavičienė [3], they pointed out that the Fourier analysis may not be advantageous for investors forecasting stock market prices as it fails to detect existing predominant cycles. An attempt to identify significant periods in the US stock market data using FFT, a method of Fourier analysis, proved to be unacceptable. Similar failures can be expected with other liquid investment instruments or financial data series. Despite this, Fourier analysis is still used for forecasting in finance and its benefits are a topic of discussion among financial market practitioners and academicians.

III. MATERIALS

Placeholder line

A. DATASET

The dataset comprises historical daily closing stock prices (in Vietnamese Dong - VND) for three prominent Vietnamese real estate companies:

- Quoc Cuong Gia Lai Joint Stock Company (QCG)
- Dat Xanh Group Joint Stock Company (DXG)
- Vinhomes Joint Stock Company (VHM)

The data spans a five-year period from March 1, 2019, to March 1, 2024. While the raw data includes additional attributes such as opening price, high, low, volume, and change, this study focuses solely on the "Close" price to develop predictive models for future closing price movements.

B. DESCRIPTIVE STATISTICS

TABLE 1. QCG, VHM, DXG's Descriptive Statistics

	DXG	VHM	QCG
Observations	1252	0	1252
Mean	17676	0	7586.17
Median	15348	0	7105
Std	7862	0	3102.55
Min	6739	0	3320
Max	46750	0	23200
25%	12303	0	4960
50%	15348	0	7105
75%	20806	0	9182.5
Skewness	1.41	0	1.13
Kurtosis	1.97	0	1.68

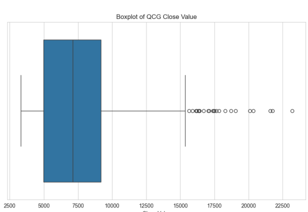

FIGURE 1. QCG stock price's boxplot

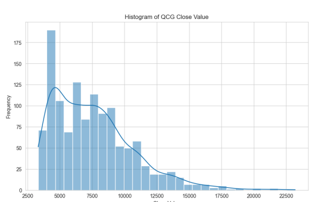

FIGURE 2. QCG stock price's histogram

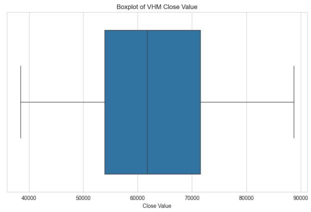

FIGURE 3. VHM stock price's boxplot

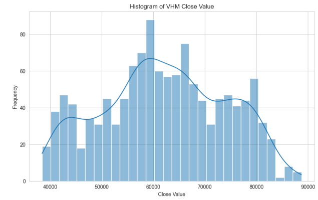

FIGURE 4. VHM stock price's histogram

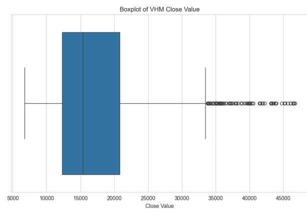

FIGURE 5. DXG stock price's boxplot

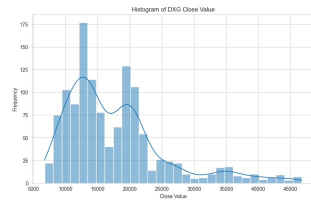

FIGURE 6. DXG stock price's histogram

IV. METHODOLOGY

Placeholder line

A. LINEAR REGRESSION

Placeholder

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

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.
- $\beta_1, \ldots, \beta_k$ are the regression coefficients for the independent variables.
- ε is the error term.

B. LONG SHORT TERM MEMORY (LSTM)

Long Short Term Memory networks (LSTM), often known as LSTMs, are a special type of recurrent neural network (RNN) with the ability to learn and remember long-term dependencies. LSTMs were introduced by Hochreiter and Schmidhuber in 1997, and have since been refined and developed further by many researchers and experts in the field. Thanks to their exceptional performance on various tasks, LSTMs have become increasingly popular.

LSTMs are designed to address the problem of long-term dependencies. Retaining information over extended periods is an inherent characteristic of LSTMs, requiring no special training to achieve this capability. In other words, the ability to remember long-term information is built into LSTMs.

Unlike traditional RNNs, which have a simple structure with a single tanh activation layer, LSTMs have a more complex chain-like structure, with modules that contain up to four layers interacting in a special way.

In the t-th state of the LSTM model:

Output: c_t, h_t, where c is the cell state, and h is the hidden state.

Input: c_{t-1}, h_{t-1}, x_t, where x_t is the input at state t of the model, and c_{t-1} and h_{t-1} are the outputs from the previous layer. The hidden state h is similar to s in RNN, while c is the unique aspect of LSTM.

Reading the diagram: The symbols σ and tanh indicate that the step uses the sigmoid and tanh activation functions, respectively. The multiplication is element-wise, and the addition is matrix addition.

Gates: f_t, i_t, o_t correspond to the forget gate, input gate, and output gate, respectively.

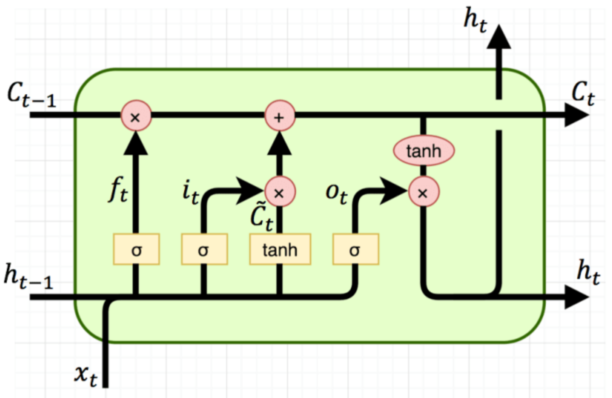

FIGURE 7. LSTM Model

- **Forget gate**:

$$f_t = \sigma(U_f \cdot x_t + W_f \cdot h_{t-1} + b_f)$$

- **Input gate**:

$$i_t = \sigma(U_i \cdot x_t + W_i \cdot h_{t-1} + b_i)$$

- **Output gate**:

$$o_t = \sigma(U_o \cdot x_t + W_o \cdot h_{t-1} + b_o)$$

Thus, the expressions for each gate of the LSTM illustrate how each gate manages the information flowing in and out of the model's states.

C. TIMESNET

Placeholder

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

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.

V. RESULT

Placeholder line

A. EVALUATION METHODS

Mean Percentage Absolute Error (MAPE): is the average percentage error in a set of predicted values.

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

Root Mean Squared Error (RMSE): is the square root of average value of squared error in a set of predicted values.

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

Mean Absolute Error (MSLE):is the relative difference between the log-transformed actual and predicted values.

$$MSLE = \frac{1}{n} \sum_{i=1}^{n} (log(1 + \hat{y}_i) - log(log(1 + y_i))^2$$

Where:

- n is the number of observations in the dataset.
- y_i is the true value.
- $\hat{y}_i$ is the predicted value.

B. DXG DATASET

TABLE 2. DXG Dataset's Evaluation

C. VHM DATASET

D. QCG DATASET

VI. CONCLUSION

Placeholder line

A. SUMMARY

Placeholder

B. FUTURE CONSIDERATIONS

Placeholder line

ACKNOWLEDGMENT

Placeholder line

REFERENCES

[1] Hind Daori, Alanoud Alanazi, Manar Alharthi, Ghaida Alzahrani , "Predicting Stock Prices Using the Random ForestClassifier", November, 14th, 2022.
[2] Hugo Souto, 2023, "TimesNet for Realized Volatility Prediction ".
[3] Bohumil Stádník, Jurgita Raudeliuniene, Vida Davidavičienė , 2016. Fourier Analysis For Stock Price Forecasting: Assumption And Evidence .