International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 2, April 2024, pp. 1906~1915
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i2.pp1906-1915 ❒ 1906

Feature selection using non-parametric correlations and important features on recursive feature elimination for stock price prediction

Arif Mudi Priyatno[1], Wahyu Febri Ramadhan Sudirman[2], Raja Joko Musridho[3]
[1]Department of Digital Business, Faculty of Economics and Business, Pahlawan Tuanku Tambusai University, Riau, Indonesia
[2]Department of Islamic Banking, Faculty of Islamic Religion, Pahlawan Tuanku Tambusai University, Riau, Indonesia
[3]Department of Informatics Engineering, Faculty of Engineering, Pahlawan Tuanku Tambusai University, Riau, Indonesia

Article Info

Article history:

Received Oct 1, 2023
Revised Dec 4, 2023
Accepted Dec 18, 2023

Keywords:

Feature selection
Important features
Non-parametric correlations
Recursive feature elimination
Stock price prediction

ABSTRACT

Stock price prediction using machine learning is a rapidly growing area of research. However, the large number of features that can be used can complicate the learning process. The feature selection method that can be used to overcome this problem is recursive feature elimination. Standard recursive feature elimination carries the risk of producing inaccurate algorithms because the top-ranked features are not necessarily the most important features. This research proposes a feature selection method that combines important features and nonparametric correlation in recursive feature elimination for stock price prediction. The data features used are technical indicators and stock price history. The recursive feature elimination method is modified with important features and nonparametric correlation features. The strategy for combining important features and non-parametric features is average weight, 25:75% weight, 75:25% weight, maximum weight, and minimum weight. The performance evaluation results show that the proposed feature selection method succeeded in obtaining small error values. The proposed method for predicting PT Bank Rakyat Indonesia Tbk (BBRI) stock prices obtains mean squared error, root mean square error, mean absolute error, and mean absolute percentage error evaluation values of 0.0000336, 0.00577, 0.00459, and 1.78%, respectively. This shows that recursive feature elimination with feature selection that combines important features and non-parametric correlation works better than the original recursive feature elimination at predicting stock prices.

Corresponding Author:

Arif Mudi Priyatno
Department of Digital business, Faculty of Economics and Business, Pahlawan Tuanku Tambusai University
Riau, Indonesia
Email: arifmudi@universitaspahlawan.ac.id

1. INTRODUCTION

People from various countries are very interested in the stock market as an alternative to investing. Investing in the stock market is a promising way to get large profits in a relatively short time, but it also has high risks [1]. As a result, information that is accurate and trustworthy must support the decision to invest in the stock market. To get accurate information about stock price movements, one of the commonly used approaches is technical analysis or fundamental analysis [2]. Technical analysis involves the use of indicators to predict stock price movements [3], while fundamental analysis involves evaluating company performance, government policies, and other economic factors [4]. Both analyzers can be used simultaneously or not. This analysis is tailored to investors. Long-term investors pay attention to fundamental analysis, while short-term and medium-term investors pay attention to technical analysis.

Although these two approaches can provide useful information for investors, limitations in manual analysis methods can result in less accurate predictions. This is because there are other factors that can influence stock prices, such as political, social, and security factors. Therefore, researchers and market players have looked for alternative methods to increase the accuracy of stock price predictions. This decade, there has been a lot of stock price prediction research using machine learning. Shah *et al.* [5] predicts stock prices by utilizing Twitter sentiment and machine learning. Subasi *et al.* [6] uses machine learning methods such as artificial neural networks, K-nearest neighbor, support vector machines, and random forests to predict stock prices. Khan *et al.* [7] uses machine learning methods naive Bayes, sequential minimal optimization, k-nearest neighbor, locally weighted learning, attribute selected classifier, partial C4.5 decision tree (PART), multilayer perceptron, random forest, bagging, and decision tree for predicting stock price movement trends. Sawale and Rawat [8] uses machine learning methods, such as artificial neural networks and support vector machines, to predict stock prices. Pahlawan *et al.* [9] carried out stock price forecasting by utilizing macroeconomics and recurrent neural network machine learning. Nam *et al.* [10] carried out on furcating stock prices by utilizing big data. Fattah *et al.* [11] predicts stock price trends using auto-machine learning.

These studies have shown that machine learning can predict stock prices more accurately than traditional methods, such as technical and fundamental analysis. However, there are several challenges in applying machine learning to predict stock prices, namely data dimensions, noise, and redundant data. In addition, the large number of features from technical indicators and price history can make it difficult for machine learning to learn price patterns because irrelevant and uncorrelated features can interfere with the learning process. Feature selection methods have been widely used to overcome this problem. One feature selection method that can be used in the stock case is recursive feature elimination [12]–[15]. Xu *et al.* [12] carried out stock price trends with support vector machine recursive feature elimination and random forest recursive feature selection. The results of support vector machine recursive feature elimination are better than those of random forest recursive feature selection. Botunac *et al.* [15] predicts stock prices using recursive feature elimination and a long short-term memory neural network. The results show that feature selection has a significant impact on stock price prediction using long short-term memory neural network. Nagaraj *et al.* [13] carried out a comparison of stock price predictions using artificial neural networks and random forests, as well as recursive feature elimination and feature selection on technical indicators. The results show that random forest with recursive feature elimination and feature selection gets the best results. However, the recursive feature elimination method has a significant weakness, namely that it has the risk of producing an inaccurate algorithm. This is because this method of getting the top-ranking features is not necessarily the most important feature.

In this research, we proposed feature selection using important features and nonparametric correlation in recursive feature elimination for stock price prediction. The data features used are technical indicators and stock price history. The recursive feature elimination method is modified with important features and non-parametric correlation features are used to reduce the number of features used. Stock price prediction using machine learning.

This article is structured in this way: in section 2, the research methods are described. With a framework image, it will explain the research. The first stage of data collection. The second stage is forming technical indicators. In the third stage of feature selection, the standard recursive feature elimination method and the proposed recursive feature elimination method are explained. The fourth stage of the prediction method is used, and the fifth stage of evaluation is used. Section 3 presents the results and discussion of the research obtained. This section allows careful examination and comparison of the results. The article concludes in section 4 with a summary of the main conclusions drawn from the research.

2. METHOD

Figure 1 is the framework for this research. In general, the stages of this research are stock data collection, the formation of technical indicators, feature selection with recursive feature elimination, prediction using machine learning, and performance evaluation. Stock data is taken from the Indonesian stock exchange. Recursive feature elimination feature selection is used, namely modification and original. The machine learning methods used are linear regression (LR), support vector regression (SVR), long short-term memory (LSTM), bidirectional LSTM (Bi-LSTM), and gated recurrent unit (GRU). The performance evaluation used is mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

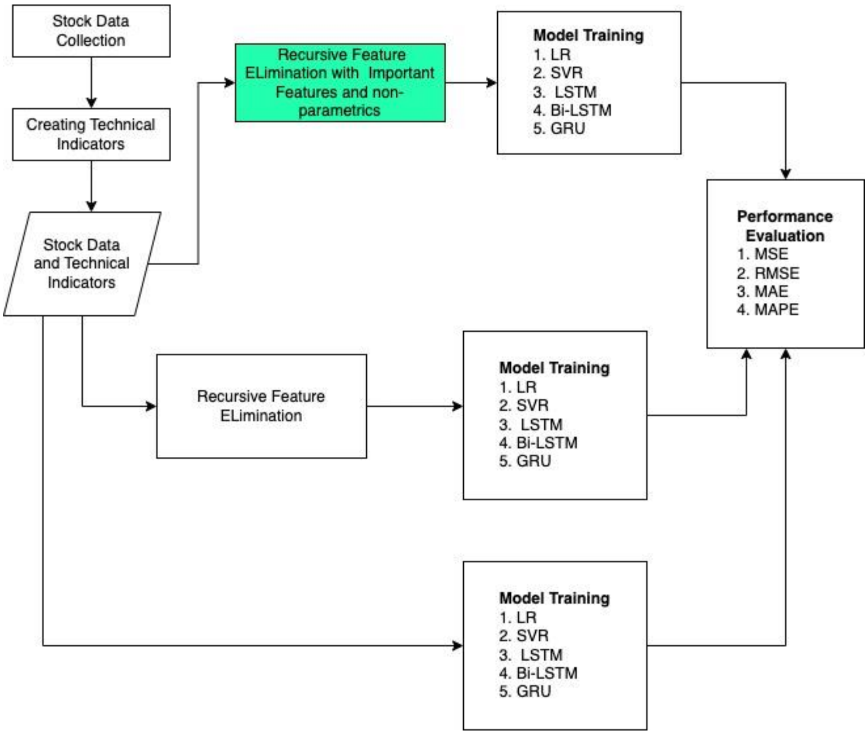

Figure 1. Framework for this research

2.1. Data collection

The stock data used in this research are stock of PT Bank Central Asia Tbk (BBCA), PT Bank Negara Indonesia (Persero) Tbk (BBNI), PT Bank Rakyat Indonesia Tbk (BBRI), and PT Bank Mandiri Tbk (BMRI). Stock history data was taken from the Indonesian stock exchange (IDX). The first initial public offering of BBCA, BBNI, BBRI, and BMRI was in 2000, 1996, 2003, and 2003, respectively. Historical data was taken from January 2, 2015, until March 6, 2023. The stock data columns taken are date, previous, *open_price*, *first_trade*, *high*, *low*, *close*, *change*, *volume*, *value*, *frequency*, *index_individual*, *offer*, *offer_volume*, *bid*, *bid_volume*, *listed_shares*, *tradeable_shares*, *weight_for_index*, *foreign_sell*, *foreign_buy*, *delisting_date*, *non_regular_volume*, *non_regular_value*, and *non_regular_frequency*.

2.2. Technical indicators

Technical indicators are tools used to analyze asset price movements, such as stock. Technical indicators use historical price data to predict future price movements. The technical indicators used in this research are momentum and volatility. The list of technical indicators based on momentum used is simple moving average, exponential moving average, weight moving average, relative strength index, money flow index, Williams indicator, momentum, accelerator oscillator, balance of power, and real moving average. The list of technical indicators based on volatility used is: true range, average true range, Bollinger bands, triple exponentials average, average directional movement index, and arc oscillator [16]–[21].

2.3. Feature selection

Feature selection in this research uses modified recursive feature elimination. Algorithm 1 is a modified recursive feature elimination feature selection using important features and non-parametric correlation. Important features are obtained from the impurity function in random forest regression and non-parametric correlation is obtained from Kendal's tau correlation. The technique for combining important features and non-parametric correlation is 5 approaches. This approach means (1), 25 important feature weights and 75 non-parametric correlation weights (2), 75 important feature weights and 25 non-parametric correlation weights (3), maximum. (4), and minimum (5).

$$F_{means} = \frac{F2+F1}{2} \tag{1}$$

$$F_{2575} = 25 * F2 + 75 * F1 \tag{2}$$

$$F_{7525} = 75 * F2 + 25 * F1 \quad (3)$$

$$F_{max} = max\ (F2, F1) \quad (4)$$

$$F_{min} = min\ (F2, F1) \quad (5)$$

Algorithm 1. Non-parametric correlation and feature importance in recursive feature elimination algorithms
Input: Features the result of combining features, Number of desired features, Random forest regressor parameters
Output: List of relevant features for stock price prediction
Step:
1. Enter all the features resulting from combining the features.
2. Enter the number of features desired.
3. Enter the random forest regressor parameters.
4. Calculate non-parametric feature correlations using Kendall's tau (F1)
5. The training process uses random forest regressors.
6. Calculate the important features of random forest regressor training results (F2)
7. Combine features from non-parametric correlation calculations (F1) and important features from random forest regressors (F2).
8. Sort the merged features (F) from highest to lowest.
9. Eliminate features with the lowest impact.
10. Repeat steps 4 to 9 until all features have been eliminated and the order of feature influence is obtained.
11. Get influential features according to the desired amount.
12. Important features that influence stock price predictions are stored in the matrix.

2.4. Prediction

Data partitioning between training data and test data uses time-series cross-validation. The time series cross-validation (CV) used is 5, with 30 days of test data for each cross-validation. Stock price predictions in this research use several machine learning methods. These methods include linear regression [22], [23], support vector regression (SVR) [24], long short-term memory networks (LSTM) [15], bi-directional LSTM [25], and gated recurrent units [26]. This research compares with research on recursive feature elimination (RFE) and LSTM [15], linear regression [23], support vector regression (SVR) [24], and bi-directional LSTM [25].

2.5. Performance evaluation

Performance evaluation uses mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). MSE is a good measure to assess the performance of a regression model because it can indicate how accurate the model is in predicting numerical values. A low MSE indicates that the regression model is quite accurate in predicting numerical values. A high MSE indicates that the regression model is not very accurate in predicting numerical values [27]. MSE can be calculated using (6), namely by adding the squared differences between the predicted value and the actual value, then dividing it by the amount of data [28]. RMSE is a good measure to assess the performance of a regression model because it can show how far the predicted values are from the actual values. A low RMSE indicates that the regression model is quite accurate in predicting numerical values. A high RMSE indicates that the regression model is not very accurate in predicting numerical values [23]. RMSE can be calculated using (7).

MAE is a measure of how far the predicted value is from the actual value by calculating the absolute value of the difference between the two and then taking the average [29]. To calculate MAE, we can use (8) [30]. MAPE is a good measure to assess the performance of a regression model because it can show how much error the average prediction has compared to the actual value. A low MAPE indicates that the regression model is quite accurate in predicting numerical values. A high MAPE indicates that the regression model is not very accurate in predicting numerical values [31]. How to calculate MAPE using (9) [32]. The notation *n* is the total of the data. Y_i is the actual value of the *i*-th data and Y_i is the resulting value of the *i*-th data prediction.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \quad (7)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad (8)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| x100 \quad (9)$$

3. RESULTS AND DISCUSSION

The historical number of BBCA, BBNI, BBRI, and BMRI stock is 1994 per stock. The historical stock data columns obtained are *date*, *previous*, *open_price*, *first_trade*, *high*, *low*, *close*, *change*, *volume*, *value*, *frequency*, *index_individual*, *offer*, *offer_volume*, *bid*, *bid_volume*, *listed_shares*, *tradeable_shares*, *weight_for_index*, *foreign_sell*, *foreign_buy*, *delisting_date*, *non_regular_volume*, *non_regular_value*, and *non_regular_frequency*. The columns used from stock history are *date*, *previous*, *open*, *high*, *low*, *close*, *change*, *volume*, *value*, *frequency*, *bid*, and *offer*. Stock has the right to a stock-split corporate action. However, BBCA stock carried out corporate actions in the data collection range. The BBCA stock split corporate action was carried out on October 13, 2021. The stock split ratio is 1:5. Table 1 is stock price data obtained from IDX. Therefore, BBCA stock is adjusted according to the stock split ratio.

Table 2 is the result of adjusting the stock split results according to the ratio. The latest stock data in Table 2 shows a difference of 1:5 compared to Table 1. This is in accordance with the stock split ratio for BBCA stock. The formation of technical indicators uses stock data obtained from IDX for BBRI, BMRI, and BBNI, while BBCA uses data after stock split adjustments. The setting for the formation of technical indicators in this research is 3–75 days. The total number of indicator technical result features is 1,152 indicator technical features.

Table 3 features the results of the feature selection process with the first 100 features from the original modification and recursive feature elimination feature selection. The results of the proposed feature selection method and the original recursive feature elimination on BBNI and BMRI stock on the first feature are different. The BBNI results with the original RFE get the close feature on that day, while the proposed method gets the previous feature, which is the most important for stock predictions for the next day. This will make the prediction results immediately different. In features 98, and 100, the proposed method and the original recursive feature elimination method get different features.

Table 1. PT Bank Central Asia Tbk (BBCA) stock price

Date	Previous	Open	First Trade	High	Low	Close	Change
2015-01-02	13,125	13,275	13,275	13,275	13,159	13,225	100
2015-01-05	13,225	13,150	13,200	13,200	13,125	13,200	-25
2015-01-06	13,200	13,000	13,050	13,200	13,000	13,100	-100
2015-01-07	13,100	13,050	13,200	13,200	13,050	13,125	25
2015-01-08	13,125	13,125	13,125	13,150	12,975	12,975	-150

Table 2. PT Bank Central Asia Tbk (BBCA) stock price after adjusting for the stock split action

Date	Previous	Open	First Trade	High	Low	Close	Change
2015-01-02	2,625	2,675	2,675	2,675	2,650	2,650	25
2015-01-05	2,650	2,650	2,650	2,650	2,625	2,650	0
2015-01-06	2,650	2,600	2,600	2,650	2,600	2,625	-25
2015-01-07	2,625	2,625	2,625	2,650	2,625	2,625	0
2015-01-08	2,625	2,625	2,625	2,650	2,600	2,600	-25

Table 3. Modified RFE feature selection results

Name	1st Feature	2nd Feature	3rd Feature	98th Feature	99th Feature	100th Feature
BBCA means	Previous	Close	Open	rma_8	rma_9	rma_10
BBCA 2575	Previous	Close	Open	rma_9	rma_10	rma_11
BBCA 7525	Previous	Close	Open	rma_8	rma_9	rma_10
BBCA max	Previous	Close	Open	rma_9	rma_10	rma_11
BBCA min	Previous	Close	Open	bop_75	rma_3	rma_74
BBNI means	Previous	Close	Open	rma_8	rma_9	rma_10
BBNI 2575	Previous	Close	Open	rma_8	rma_9	rma_10
BBNI 7525	Previous	Close	Open	rma_8	rma_9	rma_10
BBNI max	Previous	Close	Open	rma_8	rma_9	rma_10
BBNI min	Close	High	Low	bop_61	bop_64	bop_75
BBRI means	Previous	Close	Open	rma_7	rma_8	rma_9
BBRI 2575	Previous	Close	Open	rma_7	rma_8	rma_9
BBRI 7525	Previous	Close	Open	rma_7	rma_8	rma_9
BBRI max	Previous	Close	Open	rma_7	rma_8	rma_9
BBRI min	Previous	Close	Open	bop_73	bop_74	rma_58
BMRI means	Previous	Close	Open	rma_8	rma_9	rma_10
BMRI 2575	Previous	Close	Open	rma_8	rma_9	rma_10
BMRI 7525	Previous	Close	Open	rma_8	rma_9	rma_10
BMRI max	Previous	Close	Open	rma_8	rma_9	rma_10
BMRI min	Close	Open	High	bop_75	rma_3	rma_74

A machine learning-based stock price prediction process uses the data from feature selection. The machine learning used in this research is linear regression, support vector regression, long-short-term memory networks, gated recurrent units, and bidirectional LSTM. The prediction process was carried out by cross-validation five times. The experimental features carried out were 7 features, 67 features, and 100 features. The evaluations used are mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

Figures 2 to 5 show the test evaluation results using four error measures (MSE, RMSE, MAE, and MAPE), with a total of seven features resulting from the RFE combination. This test evaluation was carried out to determine the performance of the stock price prediction model with various error sizes. The MSE test with seven features shown in Figure 2 found that the best value was found in predicting the price of BBRI stock using the 2575 features combination method and linear regression machine learning. The MSE value obtained is 0.0000336. In Figure 3, the RMSE evaluation with seven features gets the best value using the approach of combining 2575 features and BBRI stock as well as linear regression machine learning. The best RMSE value is 0.00577. In Figure 4, the MAE evaluation with seven features gets the best value of 0.00459. This MAE value was obtained using the approach of combining 7525 features and BBRI stock as well as linear regression machine learning. In Figure 5, MAPE evaluation with seven features gets the best value of 1.78%. The maximum pooling approach on BBCA stock and machine learning with linear regression helped to obtain this best MAPE value. The low evaluation result values (MSE, RMSE, MAE, and MAPE) indicate that the prediction model can produce predictions that are close to the actual value.

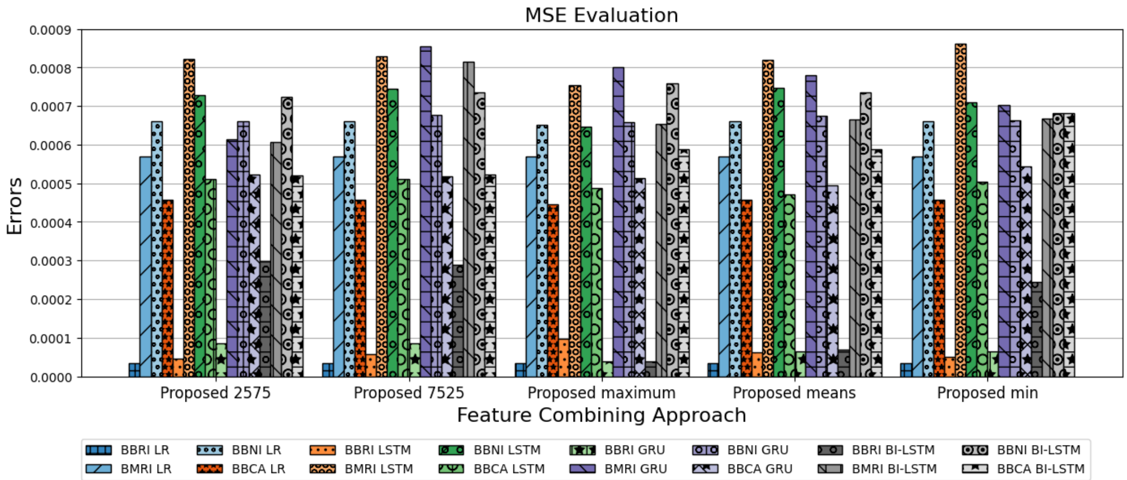

Figure 2. Evaluation using MSE with 7 features

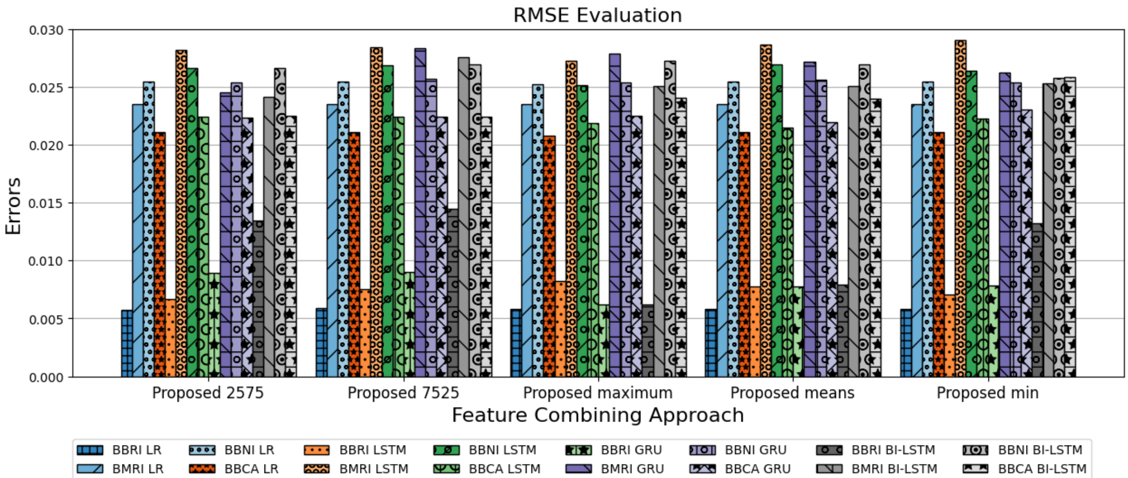

Figure 3. Evaluation using RMSE with 7 features

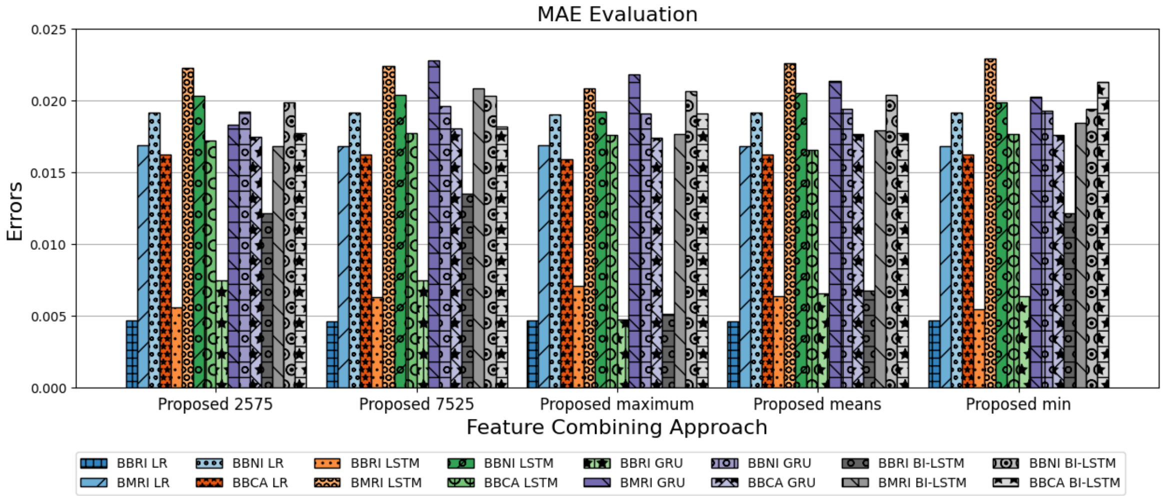

Figure 4. Evaluation using MAE with 7 features

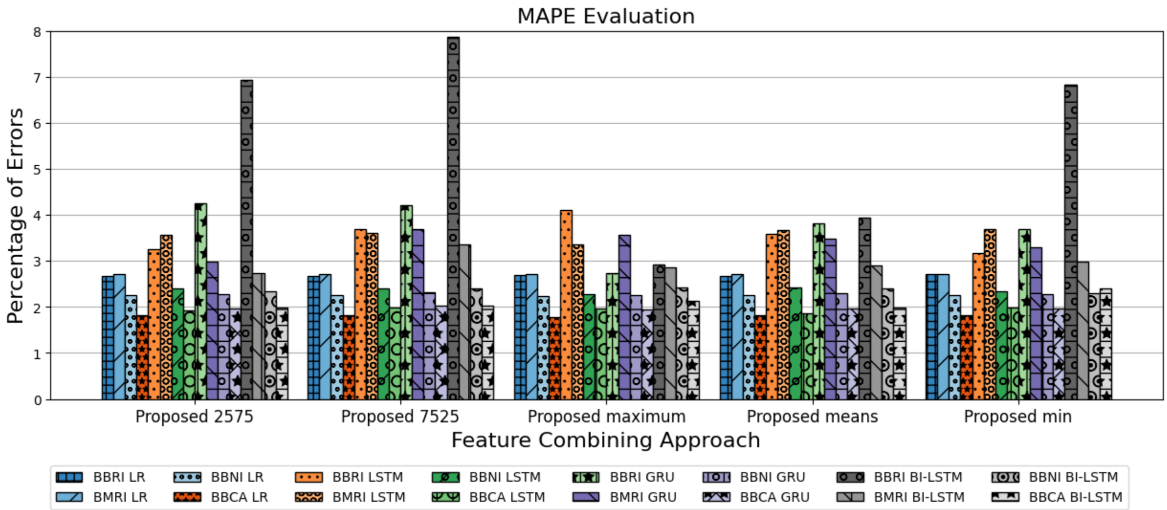

Figure 5. Evaluation using MAPE with 7 features

The results of this evaluation were compared with previous research, such as [15], [23]–[25]. The comparison results with four evaluations (MSE, RMSE, MAE, and MAPE) can be seen in Table 4. The proposed method uses seven features and a linear regression prediction method. Based on MSE, the results of the proposed method for BBRI, BMRI, BBNI, and BBCA stock have the lowest and highest values, namely 0.0000336–0.0000348, 0.0005685–0.0005691, 0.00065–0.0006602, and 0.0004463–0.0004576. The best MSE value for comparative papers was obtained through research [15]. The MSE values obtained for BBRI, BMRI, BBNI, and BBCA stock are 0.0000761, 0.0008361, 0.0007468, and 0.0005481, respectively. The difference between the MSE of research [15] and the highest MSE of the proposed method for each stock is 0.0000413, 0.000267, 0.0000866, and 0.0000905. These results show that the MSE of the proposed method is better than previous research.

Based on RMSE, the proposed method for BBRI, BMRI, BBNI, and BBCA stock has the lowest and highest values, namely 0.00577–0.00585, 0.02346–0.02348, 0.02524–0.02544, and 0.02080–0.02106. In previous research, the smallest RMSE was found in [15]. The RMSE value for each stock was 0.00739, 0.02872, 0.02692, and 0.02329. The difference between the RMSE of research [15] and the highest RMSE proposed for each stock is 0.00154, 0.00524, 0.00148, and 0.00223. This shows that overall, the proposed RMSE is better than previous research.

Based on MAE's evaluation, the proposed method for BBRI, BMRI, BBNI, and BBCA stock has the lowest and highest results, namely 0.00459–0.00466, 0.01682, 0.01896–0.01911, and 0.01590–0.01623. The closest previous research results are from [15]. The MAE results of research [15] on each stock are 0.00679, 0.02276, 0.02055, and 0.01834. The difference in MAE research [15] with the highest results proposed for each stock is 0.00213, 0.00594, 0.00144, and 0.00211. This shows that the MAE value of the proposed method is better than previous research. Based on the MAPE evaluation, the proposed method for BBRI, BMRI, BBNI, and BBCA stock has the lowest and highest results, namely 2.66% to 2.70%, 2.71% to 2.24%, and 1.78% to 1.82%. Previous research with the smallest MAPE was [15]. The MAPE values of each research stock [15] are 3.98%, 3.67%, 2.43%, and 2.06%. Based on the MAPE results, the proposed method is successful in predicting stock prices.

Table 4. Comparison proposed using linear regression with previous research

Features	BBRI	BMRI	BBNI	BBCA
		MSE		
Proposed 2575	0.0000336	0.0005685	0.0006602	0.0004572
Proposed 7525	0.0000348	0.0005691	0.0006602	0.0004573
Proposed Max	0.0000339	0.0005685	0.00065	0.0004463
Proposed Means	0.0000345	0.0005691	0.0006602	0.0004576
Proposed Min	0.0000341	0.0005691	0.0006602	0.0004573
Paper [15]	0.0000761	0.0008361	0.0007468	0.0005481
Paper [23]	0.0063072	0.0055346	0.0046166	0.0016852
Paper [24]	0.0009039	0.0038095	0.0051243	0.0194261
Paper [25]	0.0008124	0.0052381	0.0029835	0.0196963
		RMSE		
Proposed 2575	0.00577	0.02346	0.02544	0.02105
Proposed 7525	0.00585	0.02348	0.02544	0.02106
Proposed Max	0.00580	0.02346	0.02524	0.02080
Proposed Means	0.00583	0.02348	0.02544	0.02106
Proposed Min	0.00579	0.02348	0.02544	0.02106
Paper [15]	0.00739	0.02872	0.02692	0.02329
Paper [23]	0.07753	0.07352	0.06762	0.04029
Paper [24]	0.02988	0.03414	0.06850	0.13816
Paper [25]	0.02638	0.06755	0.05298	0.13064
		MAE		
Proposed 2575	0.00462	0.01682	0.01911	0.01621
Proposed 7525	0.00459	0.01682	0.01911	0.01623
Proposed Max	0.00466	0.01682	0.01896	0.01590
Proposed Means	0.00461	0.01682	0.01911	0.01623
Proposed Min	0.00466	0.01682	0.01911	0.01623
Paper [15]	0.00679	0.02276	0.02055	0.01834
Paper [23]	0.06176	0.06140	0.05503	0.03203
Paper [24]	0.02862	0.02913	0.06285	0.13487
Paper [25]	0.02513	0.06184	0.04436	0.12703
		MAPE		
Proposed 2575	2.68%	2.71%	2.26%	1.81%
Proposed 7525	2.66%	2.71%	2.26%	1.82%
Proposed Max	2.70%	2.71%	2.24%	1.78%
Proposed Means	2.67%	2.71%	2.26%	1.82%
Proposed Min	2.70%	2.71%	2.26%	1.82%
Paper [15]	3.98%	3.67%	2.43%	2.06%
Paper [23]	35.89%	10.03%	6.52%	3.61%
Paper [24]	16.64%	4.72%	7.37%	15.10%
Paper [25]	14.33%	10.09%	5.29%	14.25%

4. CONCLUSION

This research proposes feature selection using important features and nonparametric correlation in recursive feature elimination for stock price prediction. The data features used are technical indicators and stock price history. The recursive feature elimination method is modified with important features and non-parametric correlation features. The strategy for combining important features and non-parametric features is average weight, 25:75% weight, 75:25% weight, maximum weight, and minimum weight. The performance evaluation results show that the proposed feature selection method succeeds in making better predictions. The best MSE, RMSE, MAE, and MAPE evaluation results were 0.0000336, 0.00577, 0.00459, and 1.78%, respectively. The best merging approach is based on MSE and RMSE, MAE, and MAPE, respectively, namely 25 important feature weights and 75 non-parametric correlation weights, 75 important feature weights and 25 non-parametric correlation weights, and maximum weight. This shows that feature selection that combines important features and non-parametric correlation in recursive feature elimination produces stock price predictions that are more effective for predicting stock prices.

ACKNOWLEDGEMENTS

Thank you to the Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology, for financial support through the 2023 *Penelitian Dosen Pemula* (PDP) scheme with contract numbers 074/E5/PG.02.00.PM/2023, 024/LL10/PG.AK/2023, 119/LPPM/KPD/DPRM/VIII/2023. Thank you to the Institute for Research and Community Service at Pahlawan Tuanku Tambusai University for providing guidance for novice lecturers in implementing the *Penelitian Dosen Pemula* scheme.

REFERENCES

[1] M. Ratchagit and H. Xu, “A two-delay combination model for stock price prediction,” *Mathematics*, vol. 10, no. 19, Sep. 2022, doi: 10.3390/math10193447.
[2] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “A systematic review of fundamental and technical analysis of stock market predictions,” *Artificial Intelligence Review*, vol. 53, no. 4, pp. 3007–3057, Apr. 2020, doi: 10.1007/s10462-019-09754-z.
[3] J. B. Pandya and U. K. Jaliya, “Opinion and technical indicator based optimized deep learning for prediction of stock market,” *Indian Journal of Computer Science and Engineering*, vol. 12, no. 6, pp. 1860–1874, Dec. 2021, doi: 10.21817/indjcse/2021/v12i6/211206166.
[4] G. Ji, J. Yu, K. Hu, J. Xie, and X. Ji, “An adaptive feature selection schema using improved technical indicators for predicting stock price movements,” *Expert Systems with Applications*, vol. 200, Aug. 2022, doi: 10.1016/j.eswa.2022.116941.
[5] U. Shah, B. Karani, J. Shah, and M. Dhande, “Stock market prediction using sentimental analysis and machine learning,” in *2021 2nd Global Conference for Advancement in Technology (GCAT)*, Oct. 2021, pp. 1–4, doi: 10.1109/GCAT52182.2021.9587898.
[6] A. Subasi, F. Amir, K. Bagedo, A. Shams, and A. Sarirete, “Stock market prediction using machine learning,” *Procedia Computer Science*, vol. 194, no. 4, pp. 173–179, 2021, doi: 10.1016/j.procs.2021.10.071.
[7] W. Khan, U. Malik, M. A. Ghazanfar, M. A. Azam, K. H. Alyoubi, and A. S. Alfakeeh, “Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis,” *Soft Computing*, vol. 24, no. 15, pp. 11019–11043, Aug. 2020, doi: 10.1007/s00500-019-04347-y.
[8] G. J. Sawale and M. K. Rawat, “Stock market prediction using sentiment analysis and machine learning approach,” in *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Jan. 2022, pp. 1–6, doi: 10.1109/ICSSIT53264.2022.9716326.
[9] M. Reza Pahlawan, E. Riksakomara, R. Tyasnurita, A. Muklason, F. Mahananto, and R. A. Vinarti, “Stock price forecast of macro-economic factor using recurrent neural network,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, pp. 74–83, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp74-83.
[10] S. T. Nam, C. Y. Jin, and S. Y. Shin, “A forecasting of stock trading price using time series information based on big data,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 3, pp. 2548–2554, 2021, doi: 10.11591/ijece.v11i3.pp2548-2554.
[11] D. A. Fattah, A. A. Naim, A. S. Desuky, and M. S. Zaki, “AutoKeras and particle swarm optimization to predict the price trend of stock exchange,” *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 11, no. 2, pp. 1100–1109, 2022, doi: 10.11591/eei.v11i2.3373.
[12] Y. Xu, Z. Li, and L. Luo, “A study on feature selection for trend prediction of stock trading price,” in *2013 International Conference on Computational and Information Sciences*, Jun. 2013, no. 2, pp. 579–582, doi: 10.1109/ICCIS.2013.160.
[13] N. Nagaraj, B. M. Vikranth, and N. Yogesh, “Recursive feature elimination technique for technical indicators selection,” in *Communications in Computer and Information Science*, vol. 1589, A. Bennour, T. Ensari, Y. Kessentini, and S. Eom, Eds. Cham: Springer International Publishing, 2022, pp. 139–145.
[14] H. H. Htun, M. Biehl, and N. Petkov, “Survey of feature selection and extraction techniques for stock market prediction,” *Financial Innovation*, vol. 9, no. 1, Jan. 2023, doi: 10.1186/s40854-022-00441-7.
[15] I. Botunac, A. Panjkota, and M. Matetic, “The effect of feature selection on the performance of long short-term memory neural network in stock market predictions,” in *Annals of DAAAM and Proceedings of the International DAAAM Symposium*, vol. 31, no. 1, 2020, pp. 0592–0598.
[16] Y. Ayyappa and A. P. Siva Kumar, “Optimized long short-term memory-based stock price prediction with sentiment score,” *Social Network Analysis and Mining*, vol. 13, no. 1, Dec. 2023, doi: 10.1007/s13278-022-01004-5.
[17] Y. Ma, R. Mao, Q. Lin, P. Wu, and E. Cambria, “Multi-source aggregated classification for stock price movement prediction,” *Information Fusion*, vol. 91, pp. 515–528, Mar. 2023, doi: 10.1016/j.inffus.2022.10.025.
[18] H. Kang, X. Zong, J. Wang, and H. Chen, “Binary gravity search algorithm and support vector machine for forecasting and trading stock indices,” *International Review of Economics and Finance*, vol. 84, pp. 507–526, Mar. 2023, doi: 10.1016/j.iref.2022.11.009.
[19] F. Orte, J. Mira, M. J. Sánchez, and P. Solana, “A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction,” *Research in International Business and Finance*, vol. 64, no. 2, pp. 1170–1177, Jan. 2023, doi: 10.1016/j.ribaf.2022.101829.
[20] T. Bao, B. Corgnet, N. Hanaki, Y. E. Riyanto, and J. Zhu, “Predicting the unpredictable: new experimental evidence on forecasting random walks,” *SSRN Electronic Journal*, vol. 146, Jan. 2022, doi: 10.2139/ssrn.4169343.
[21] S. Deng, X. Huang, Y. Zhu, Z. Su, Z. Fu, and T. Shimada, “Stock index direction forecasting using an explainable eXtreme gradient boosting and investor sentiments,” *North American Journal of Economics and Finance*, vol. 64, Jan. 2023, doi: 10.1016/j.najef.2022.101848.
[22] D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, “Stock market predication using a linear regression,” in *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)*, Apr. 2017, pp. 510–513, doi: 10.1109/ICECA.2017.8212716.
[23] A. M. Priyatno, L. S. Tanjung, W. F. R. Sudirman, P. Cholidhazia, P. Z. Jati, and F. I. Firmananda, “Comparison random forest regression and linear regression for forecasting BBCA stock price,” *Jurnal Teknik Industri Terintegrasi (JUTIN)*, vol. 6, no. 3, pp. 718–732, 2023, doi: 10.31004/jutin.v6i3.16933.
[24] P. Chhajer, M. Shah, and A. Kshirsagar, “The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction,” *Decision Analytics Journal*, vol. 2, Mar. 2022, doi: 10.1016/j.dajour.2021.100015.

[25] M. A. Istiake Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using LSTM and bi-directional LSTM model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, Oct. 2020, pp. 87–92, doi: 10.1109/NILES50944.2020.9257950.

[26] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *ArXiv:1412.3555*, pp. 1–9, Dec. 2014, doi: 10.48550/arXiv.1412.3555.

[27] S. Vazirani, A. Sharma, and P. Sharma, "Analysis of various machine learning algorithm and hybrid model for stock market prediction using python," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, IEEE, Oct. 2020, pp. 203–207, doi: 10.1109/ICSTCEE49637.2020.9276859.

[28] K. K. Yun, S. W. Yoon, and D. Won, "Interpretable stock price forecasting model using genetic algorithm-machine learning regressions and best feature subset selection," *Expert Systems with Applications*, vol. 213, Mar. 2023, doi: 10.1016/j.eswa.2022.118803.

[29] K. Olorunnimbe and H. Viktor, "Deep learning in the stock market—a systematic survey of practice, backtesting, and applications," *Artificial Intelligence Review*, vol. 56, no. 3, pp. 2057–2109, Mar. 2023, doi: 10.1007/s10462-022-10226-0.

[30] M. Ali, D. M. Khan, H. M. Alshanbari, and A. A.-A. H. El-Bagoury, "Prediction of complex stock market data using an improved hybrid EMD-LSTM model," *Applied Sciences*, vol. 13, no. 3, Jan. 2023, doi: 10.3390/app13031429.

[31] D. Li, Y. Li, C. Wang, M. Chen, and Q. Wu, "Forecasting carbon prices based on real-time decomposition and causal temporal convolutional networks," *Applied Energy*, vol. 331, Feb. 2023, doi: 10.1016/j.apenergy.2022.120452.

[32] J. Chen, "Analysis of bitcoin price prediction using machine learning," *Journal of Risk and Financial Management*, vol. 16, no. 1, Jan. 2023, doi: 10.3390/jrfm16010051.

BIOGRAPHIES OF AUTHORS

Arif Mudi Priyatno received the S.T. degree in informatic engineering from Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Riau, Indonesia, in 2018 and the M.Kom. degree in informatic engineering from Institut Teknologi Sepuluh Nopember, in 2020 and currently studying Doctoral in electrical and informatics engineering from Universitas Negeri Malang. Currently, he is an assistant professor at Department Digital Business, Universitas Pahlawan Tuanku Tambusai. His research interests include data mining, computer vision, and natural language processing. He can be contacted at email: arifmudi@universitaspahlawan.ac.id.

Wahyu Febri Ramadhan Sudirman received a bachelor's degree in management (SE) with a specialization in finance in (2017) from Universitas Islam Negeri Maulana Malik Ibrahim Malang and obtained a Master of Science in management (M.Sc.) in (2020) from Universitas Gadjah Mada with a specialization in Islamic finance. He has research interests in capital markets, behavioral finance, corporate governance, investment, and financial management. He is also active as a reviewer in several national journals in Indonesia and has published in reputable national journals in the field of financial management. He is also active as a reviewer in several national journals in Indonesia and has published in reputable national journals in the field of financial management. He can be contacted at e-mail: wahyu.febri.id@universitaspahlawan.ac.id.

Raja Joko Musridho received the Bachelor of Engineering degree (S.T) in informatics engineering from Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Riau, Indonesia in 2014. And Master of Philosophy (M.Phil.) degree in computer science from Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia in 2019. He is an assistant professor at Informatics Engineering Department in Universitas Pahlawan Tuanku Tambusai. He has authored some Scopus indexed conference papers, one of it was indexed in Web of Science. His research interests include the applications of artificial intelligence, mobile robotics and simultaneous localization and mapping. He can be contacted at email: rajajoko@gmail.com.