



Contents lists available at [JESA](https://jesa.aks.or.id)

Journal of Engineering and Science Application

journal homepage:
<https://jesa.aks.or.id/index.php/jesa/index>



Predictive Modeling of Unilever Indonesia's Stock Prices with Linear Regression: A RapidMiner-Based Approach

Zulfan Ependi^a, Fahmi Iqbal Firmananda^b

^{ab}Department of Digital Business, Faculty of Economics and Business, Pahlawan Tuanku Tambusai University, Riau, Indonesia

Article Info

Keywords:

Linear Regression,
Stock Prediction,
RapidMiner,
Unilever Indonesia,
Investment Analysis

ABSTRACT

Investment decision-making in the capital market requires accurate analysis to minimize risk and maximize returns. This study focuses on PT Unilever Indonesia Tbk, a leading Fast-Moving Consumer Goods (FMCG) company listed on the Indonesia Stock Exchange, to evaluate its stock price prediction using the linear regression method supported by RapidMiner software. Historical stock data were collected from January 2, 2018, to June 27, 2023, including attributes such as opening price, closing price, lowest price, highest price, and trading volume. The dataset was processed using data screening and modeling techniques to construct a linear regression model for prediction. Various scenarios with different proportions of training and testing data (70/30, 80/20, 60/40, and 90/10) were tested to analyze the impact of data distribution on model performance. The evaluation results showed that the 80% training and 20% testing scenario provided the lowest Root Mean Squared Error (RMSE) of 56.699, indicating better predictive accuracy compared to other scenarios. Nevertheless, the linear regression model still produced a relatively high error rate, with the best RMSE value suggesting limitations of this approach for complex market prediction. This study concludes that while linear regression can provide a basic framework for stock price forecasting, incorporating additional economic, fundamental, and external factors could significantly improve predictive reliability. The findings offer practical insights for investors and researchers in understanding the potential and limitations of linear regression in stock market analysis.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Zulfan Ependi
Department of Digital Business,
Faculty of Economics and Business,
Pahlawan Tuanku Tambusai University, Riau, Indonesia
Email: zulfanependi@universitaspahlawan.ac.id

1. INTRODUCTION

Investing is a crucial activity for most people to achieve their financial goals. Some people invest because it will be beneficial for the future. One well-known form of investment is investing in the capital market. The capital market serves as a meeting point between those with excess funds and those in need of funds through the buying and selling of securities. Investment instruments in the capital market include stocks and bonds.

Shares are securities issued by a company or individual as a sign of ownership. A Limited Liability Company (PT) that issues shares is called an issuer [1]. Shares are a popular form of financial investment because they can provide substantial profits. Nowadays, investing in shares doesn't require significant capital; with as little as Rp100,000, you can purchase the desired shares. Profits from shares are earned when there is a difference between the purchase price and the sale price, called capital gain [2].

Investing can provide knowledge for sound financial decisions and can reduce unwanted losses. Continuous inflation in the Indonesian economy is one reason to invest. Inflation causes prices of goods and services to rise, followed by a decline in people's purchasing power. Increases in the value of investment assets can mitigate the impact of inflation, thereby maintaining purchasing power [3]. Other benefits, besides capital gains, include dividends. Dividends represent profits earned by a company and then distributed to shareholders [4].

Investors naturally expect positive returns and profits on their invested shares. Achieving good investment returns requires an investment strategy aligned with financial goals, risk tolerance, and overall financial situation. Many analytical techniques are used, such as to predict and determine when to sell shares to achieve the desired profit. However, some investors use analytical techniques that are commonplace among investors, resulting in less than optimal investment returns [5].

PT Unilever Indonesia is a company engaged in the FMCG sector with its main offices in Rotterdam, the Netherlands, and London, England [6]. On January 11, 1982, Unilever listed its shares on the Indonesia Stock Exchange. Unilever Indonesia produces products widely used by Indonesians, such as cleaning products, personal care products, and food. Unilever Indonesia continues to develop and innovate its products to meet consumer needs. PT Unilever Indonesia's net profit in 2022 was IDR 5,364,761,000 [7]. This profit is one aspect investors use to determine whether the stock is worth purchasing.

The choice of PT Unilever Indonesia as the focus of this study is not without reason. As a leader in the Fast Moving Consumer Goods (FMCG) industry, Unilever plays a significant role in the Indonesian economy, presenting attractive investment opportunities. With consistent growth and innovative business strategies, Unilever demonstrates significant potential as a subject for stock investment analysis. Furthermore, the ever-changing dynamics of the stock market require investors to have a deep understanding of the factors influencing stock price movements. In this context, predictive analysis using linear regression methods becomes highly relevant, providing crucial insights that can support investors in making more informed, data-driven decisions.

Generally, linear regression is a statistical analysis technique used to identify patterns of relationships between variables [8]. Linear regression can also be used for purposes of understanding the data or case being studied, for control purposes, and for predicting dependent variables. The use of linear regression in this era significantly assists in decision-making based on the data studied. For example, in sales, linear regression can predict sales volume for a given period [8].

Unilever's stock price is a key factor in investment decisions. Therefore, this study will predict Unilever's stock price using linear regression analysis to help investors make informed investment decisions. Linear regression has the advantage of predicting future closing stock prices. The data used in this study is Unilever's daily stock data.

2. RESEARCH METHOD

2.1. Data Collection

The data collection process in this study is a crucial step to ensure the accuracy of the analysis. The data used is historical stock data for PT Unilever Indonesia, collected from January 2, 2018, to June 27, 2023. The primary source of data collection was id.investing.com, chosen for its reliability in providing comprehensive and up-to-date financial data.

Data selection focused on key aspects of stock analysis, such as opening price, highest price, lowest price, trading volume, and closing price. These selection criteria were based on their relevance to the research objective, which is to analyze stock price movement trends. In this context, we decided to use daily data because it allows for more detailed observation of short-term stock price fluctuations. The data is presented in Table 1, with several attributes shown in Table 1.

Table 1. Attributes in the data

Column name	Type	Information
Date	date	The date the transaction occurred.
Close	Numeric	The last price that occurred on a stock
Open	Numeric	The first share price is opened in the first trading session
Low	Numeric	The lowest price of a stock that occurred on that day
Volume	Numeric	Refers to the number of shares traded on that day

Change	Numeric	The difference between the stock's last price on the previous day and the last price on that day
--------	---------	--

The linear regression method was chosen because of its ability to model the relationship between independent variables (such as opening price and trading volume) and the dependent variable (closing price). Linear regression provides a way to estimate stock price changes based on changes in independent variables, which is very useful in predicting market trends [9].

The main strength of linear regression lies in its ability to provide predictions based on statistical analysis [10]. Through this model, we can identify the significant influence of independent variables on stock prices and estimate the magnitude of potential price changes if these variables change. Furthermore, the use of linear regression also aims to ensure repeatable and verifiable analysis. With this approach, the analysis results are not solely dependent on subjective interpretations but also on standardized statistical methods recognized in the scientific community [11].

2.2. Research Flow

1) Literature Review

Information collection is the initial step in this research. Information is obtained by reading literature related to the research topic. Through literature review, researchers can detail existing knowledge and formulate a theoretical basis for further research.

2) Dataset Collection

After establishing a theoretical foundation, the research moves to the dataset collection stage. This process involves selecting dataset sources and implementing strategies to collect the necessary datasets. The data for this study were obtained from the website id.investing.com.

3) Dataset Selection

At this stage, researchers make decisions about the datasets to be used. Dataset selection involves considerations related to the research objectives and data availability.

4) Data Screening

After the datasets are selected, the data screening process is carried out. This process examines the datasets to identify and address missing values or other errors.

5) Implementation

The implementation stage is the key step taken to generate insights from the data. The use of linear regression methods for prediction and the use of Rapidminer software are techniques used in this research.

6) Data Analysis

This process uses RapidMiner to process the data obtained from the data mining process. The "Last" attribute accounts for 90% of the training data, while the remainder is used for testing.

2.3. Rapid Miner

Rapid miner is an application or software that functions to learn about data mining for data analysis, data processing, and predictive model development and is open source [9]. Data Mining is a process used to retrieve hidden or hidden information, using historical data on data to then find patterns for decision making [10]. Rapidminer's user-friendly interface makes it easy for users to use it. When Rapidminer is run, it will display a welcome perspective. The design perspective is the working display of Rapidminer, and the result perspective will display the results of the analysis.

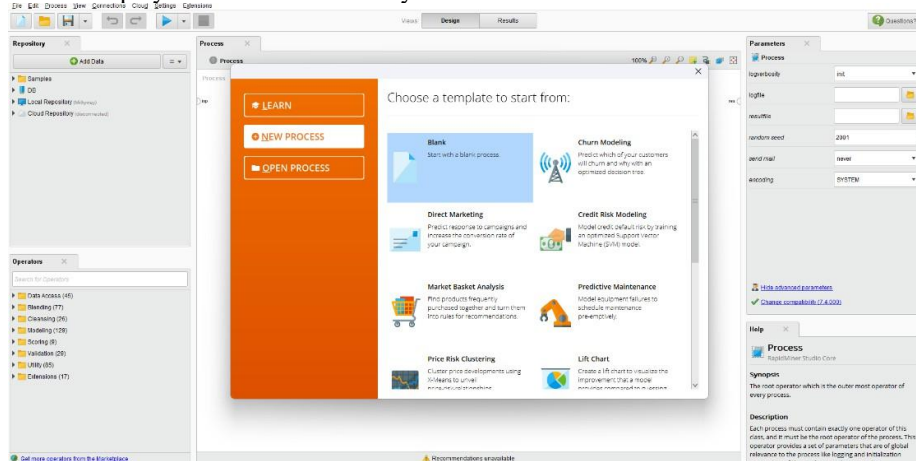


Figure 2. Initial view of Rapidminer

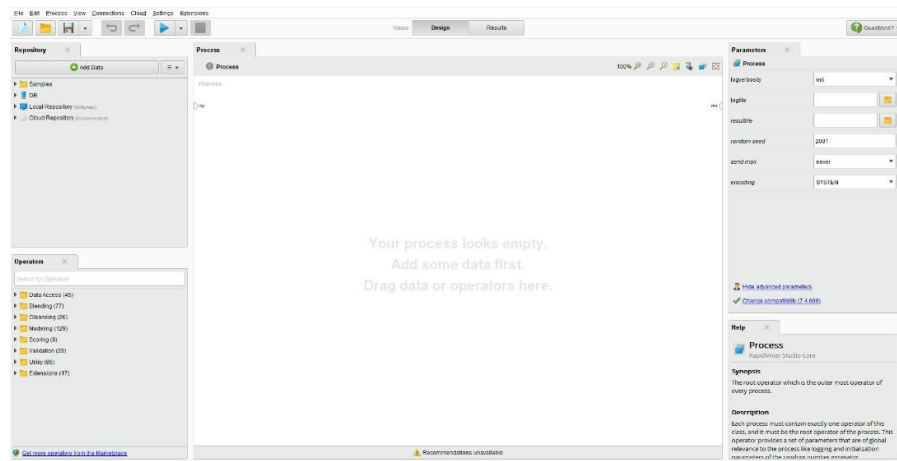


Figure 3. Design Perspective View

The design perspective view has several sections, such as view operators. In Figure 4, the view operator contains groups used in data processing. Data access contains operators for loading data files that support Excel and CSV formats, as well as cloud storage. Blending operators function to transform data such as filter, example, replace, types, and many others. Cleansing operators play a role in data preprocessing in preparation for further data analysis. Modeling operators contain models for data mining processing. Scoring operators function to evaluate the performance of models processing data that has not been previously seen. Validation operators function to test the quality and performance of models or modeling. Utilities function as additional operators such as subprocess, multiply, logging, and others. Operator extensions are used as additional modules for exporting to external formats such as Excel and database formats.

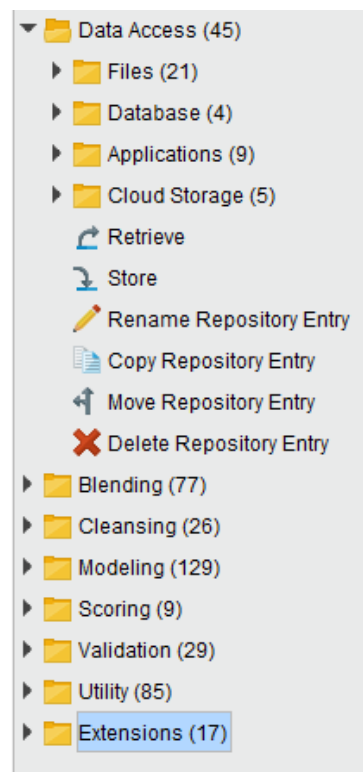


Figure 4. Rapidminer Operator

2.4. Operators used

- a) Read Excel: The Read Excel operator in Rapid Miner is used to read worksheets from Microsoft Excel. Each row in an Excel table represents a data entity, while each column represents an attribute of that data [11].
- b) Split Data: Split Data divides the desired number of data (subsets) from a given ExampleSet. This operator is very useful for machine learning data processing purposes.

- c) **Linear Regression:** The Linear Regression operator calculates a linear regression model from a given ExampleSet. Regression is used for numerical predictions. Linear regression is a statistical technique used to model linear relationships between one variable and another.
- d) **Apply Model:** Apply Model is useful for machine learning, using a model that has been trained on the latest data set. This model will use the training data to make predictions on the new data.
- e) **Performance:** The performance operator is responsible for evaluating the performance of machine learning operators. The performance operator provides a list of performance criteria such as accuracy, precision, kappa statistic, root mean squared error, mean squared error, and so on.

2.5. Linear Regression

Regression is a construction method for predicting the value of given input data. Regression analysis is a statistical method used to assess the extent of the relationship between dependent variables and independent variables. The main process in forecasting involves building a regression model by finding the dependency between one or more independent variables and the dependent variable [10]. Modeling in linear regression is to describe the relationship between a scalar variable and one or more explanatory variables. In the linear regression method, the method pattern is compiled based on past data [12]. Previously collected training data is used to form the model.

3. RESULTS AND DISCUSSION

The process begins by selecting the Read Excel operator to load the selected data. Then, click the Import Configuration Wizard and select the data to be processed. When the file is selected, the table attribute settings will appear.

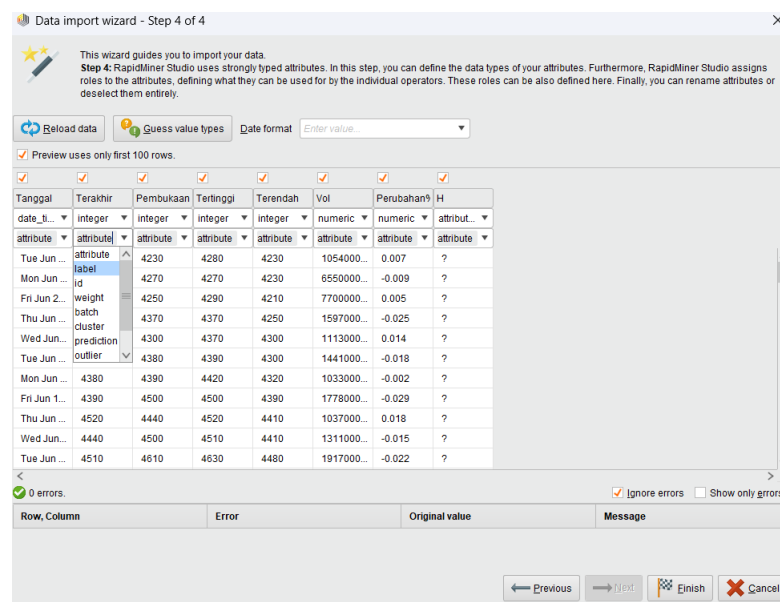


Figure 5. Changing attributes in a table

Figure 5 changes the attributes in the "Last" column because the "Last" column is the dependent variable to be predicted. Then, the Split Data operator is added. This operator is used to divide the data into two parts: training data and testing data. Training data is used to learn patterns in machine learning, while testing data is used to test the patterns that have been discovered. Then, connect the out in the Read Excel operator with the exa in the operator.

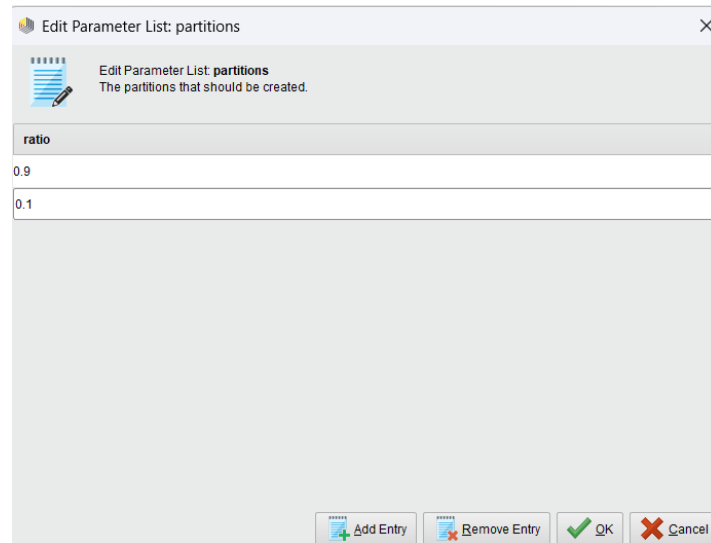


Figure 6. Dividing data into training and testing data

Then, to divide the data accordingly, select edit numeration. Figure 6 shows 90% of the training data, with the remainder being used as testing data. The next step is to add the Linear Regression operator. The line par 1 is connected to tra in the Linear Regression and mod to res, producing the linear regression results shown in Figure 7.

LinearRegression

$$0.997 * \text{Pembukaan}$$
$$- 0.000 * \text{Vol}$$
$$+ 5720.690 * \text{Perubahan\%}$$
$$+ 19.782$$

Figure 7. Results of Linear Regression

The next step is to connect the existing ports as shown in Figure 8. Line 1 on the split data output is the training data and Line 2 is the testing data.

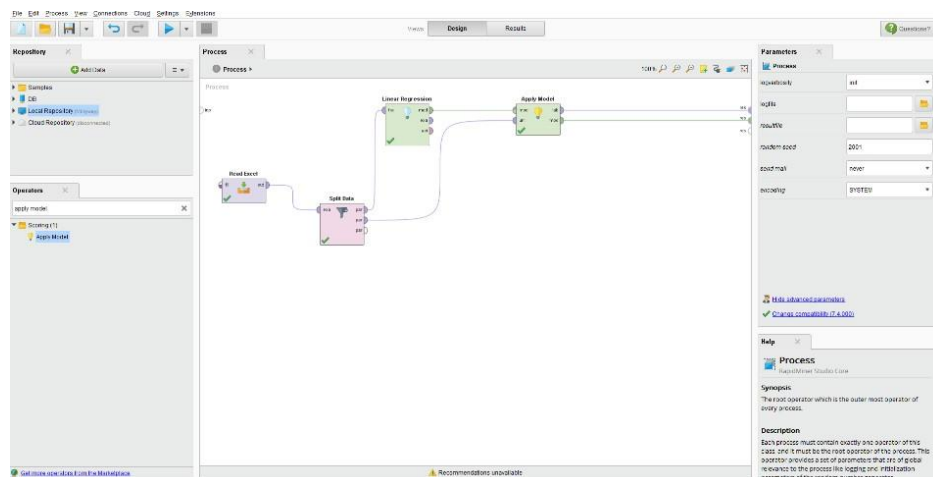


Figure 8. Performing testing on testing data

between training and testing data. Interestingly, the scenario with 90% training data and 10% testing data showed a slightly higher RMSE value (57.072) compared to the 80/20 scenario, indicating that while increasing training data generally improves accuracy, there is a point where adding more training data may not provide significant improvements or may even lead to overfitting. The data show a general trend that increasing the proportion of training data tends to improve model accuracy, but there is an optimal limit where further additions no longer produce significant improvements. In this case, the optimal scenario appears to be with 80% training data, which provides the lowest RMSE value.

Table 2. Test results by adjusting the number of training and testing data

Data Training	Data Testing	RMSE
70%	30%	64.304 +/- 0.000
80%	20%	56.699 +/- 0.000
60%	40%	61.84 - 0.000
90%	10%	57.072 +/- 0.000

Increasing the training data percentage showed a positive impact on model performance, with improved accuracy and better predictive ability on the test data. At 80% training data percentage, this was indicated by a significant decrease in the RMSE value. Therefore, the optimal percentage of 80% can be considered an effective strategy for improving the performance of linear regression models.

4. CONCLUSION

Based on the results of the research conducted with different amounts of training data, it can be concluded that the linear regression method has poor performance with a root mean squared value of 57,072. This indicates that the performance of linear regression is not very good. Other factors or variables that can improve the accuracy of the prediction value can be added other factors that affect Unilever's stock price such as; economic crisis, circulating news, the company's fundamental condition, politics, currency exchange rates, and so on.

REFERENCES

- [1] R. Christian Mahendra, "Literasi Saham pada Mahasiswa STIESIA Surabaya," JPUA J. Perpust. Univ. Airlangga Media Inf. dan Komun. Kepustakawanan, vol. 8, no. 1, p. 8, 2020, doi: 10.20473/jpua.v8i1.2018.8-13.
- [2] D. Rustiana and S. Ramadhani, "Strategi di Pasar Modal Syariah," J. Ilmu Komputer, Ekon. dan Manaj., vol. 2, no. 1, pp. 1578–1589, 2022.
- [3] S. J. L. Tobing, H. Situmorang, M. Malau, P. Y. Samosir, P. S. J. Kennedy, and A. J. Lekhenila, "Simulasi Saham dalam 'Yuk Nabung Saham' kepada Masyarakat Melalui Webinar Online," Ikraith-Abdimas, vol. 4, no. 2, pp. 143–147, 2021.
- [4] Lisa Kustina, Otika Safitri, and Samsul Anwar, "Kebijakan Deviden Dan Capital Gain: Pengaruhnya Terhadap Harga Saham," J. Investasi, vol. 5, no. 1, pp. 24–37, 2019, doi: 10.31943/investasi.v5i1.16.
- [5] A. S. Nuriyah, R. Saladini, K. N. Fujiah, and F. Indrayatna, "Prediksi Harga Saham Bank BCA Menggunakan Metode Regresi Linier Berganda," Semin. Nas. Stat. Aktuaria Ii, 2023.
- [6] D. Safitri, P. Tazkia, X. Amanda, Y. Shintia, T. F. Putri, and S. N. Amada, "Analisa Strategi Bisnis pada PT. Unilever Indonesia," J. Publ. Ilmu Manaj., vol. 2, no. 2, pp. 2963–8712, 2023.
- [7] M. P. D. M. B. Y. N. Telaumbanua, "Analisis Net Profit Margin Pada Pt Unilever Indonesia Tbk (Perusahaan Manufaktur Yang Terdaftar Pada Bursa Efek Indonesia)," J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt., vol. 10, no. Vol. 10 No. 4 (2022): JE. VOL 10 NO 4 (2022), pp. 1462–1471, 2022.
- [8] D. Kurniawan, "Regresi Linier," Statistic, pp. 1–6, 2008.
- [9] T. N. Padilah and R. I. Adam, "Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang," FIBONACCI J. Pendidik. Mat. dan Mat., vol. 5, no. 2, p. 117, 2019, doi: 10.24853/fbc.5.2.117-128.
- [10] M. Syafruddin, L. Hakim, and D. Despa, "Metode Regresi Linier Untuk Prediksi Kebutuhan Energi Listrik Jangka Panjang (Studi Kasus Provinsi Lampung)," J. Inform. dan Tek. Elektro Terap., vol. 2, no. 2, 2014, doi: 10.23960/jitet.v2i2.237.
- [11] S. Ningsih and H. H. Dukalang, "Penerapan Metode Suksesif Interval pada Analsis Regresi Linier Berganda," Jambura J. Math., vol. 1, no. 1, pp. 43–53, 2019, doi: 10.34312/jjom.v1i1.1742.
- [12] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," J. Nas. Teknol. dan Sist. Inf., vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
- [13] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," JURIKOM (Jurnal Ris. Komputer), vol. 7, no. 1, p. 56, 2020, doi: 10.30865/jurikom.v7i1.1947.
- [14] F. Akthar, C. Hahne, and O. Reference, "RapidMiner 5," p. 990, 2012, [Online]. Available: <http://www.rapid-i.com>
- [15] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, "Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi," J. Nas. Tek. Elektro dan Teknol. Inf., vol. 8, no. 4, p. 364, 2019, doi: 10.22146/jnteti.v8i4.537.
- [16] Priyatno, Arif Mudi, et al. "Coefficient-Average Based Recursive Feature Elimination Modification to Optimize Feature Selection in Stock Price Prediction." International Journal of Intelligent Engineering & Systems 18.6 (2025).
- [17] Priyatno, Arif Mudi, Didik Dwi Prasetya, and Triyanna Widiyaningtyas. "Optimizing Feature Selection for Stock Price Prediction Using a Two-Stage Feature Selection Framework." 2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C). IEEE, 2024.

- [18] Priyatno, Arif Mudi, Wahyu Febri Ramadhan Sudirman, and Raja Joko Musridho. "Feature selection using non-parametric correlations and important features on recursive feature elimination for stock price prediction." *International Journal of Electrical & Computer Engineering (2088-8708)* 14.2 (2024).
- [19] Priyatno, Arif Mudi, Lidya Ningsih, and Muhammad Noor. "Harnessing machine learning for stock price prediction with random forest and simple moving average techniques." *Journal of Engineering and Science Application* 1.1 (2024): 1-8.
- [20] Priyatno, Arif Mudi, et al. "Impurity-Based Important Features for feature selection in Recursive Feature Elimination for Stock Price Forecasting." *J. Tek. Ind. Terintegrasi* 6.4 (2023): 1182-1194.