International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 1, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

# Stock Price Prediction Model for KLBF using Linear Regression Algorithm

M. Makmun Effendi[1], Ahmad Turmuzi Zy[2], Isarianto[3]

**Abstract**
Stock prices are influenced by constantly changing supply and demand, leading to short-term price fluctuations. Other factors affecting stock prices include interest rates, inflation, company earnings, and marketing strategies. These price fluctuations increase the losses, making stock price predictions crucial to assist investors in making safer investment decisions. This research utilizes the Linear Regression algorithm to predict the stock prices of the pharmaceutical company with the stock code KLBF using a time series dataset in the period from January 2020 to January 2022. According to the experimental result, the proposed model can produce the total sum of squared errors is 27,105 with RMSE = 23.06. This low error margin indicates a strong predictive performance and the effectiveness of the proposed approach in predicting stock prices.

***Keywords:***
Prediction, Linear Regression, RMSE, KLBF



## 1. Introduction

In recent years, stock price prediction has become a crucial topic in finance due to the increasing complexity and volatility of global financial markets. Accurate forecasting of stock prices allows investors and institutions to optimize their portfolios, mitigate risks, and improve decision-making. Despite numerous advancements in machine learning and artificial intelligence, traditional statistical methods, particularly linear regression, continue to be widely used because of their simplicity, interpretability, and computational efficiency. Linear regression models the relationship between dependent variables such as stock prices that have one or more independent variables, providing fundamental insights into market behavior [1][2].

One reason linear regression remains relevant is that many stock price movements are influenced by quantifiable financial and economic indicators that naturally fit into a linear modeling framework. For example, factors like earnings per share (EPS), price-to-earnings ratio (PER), return on equity (ROE), and macroeconomic variables often display approximate linear relationships with stock prices in the short term. This makes linear regression a natural starting point for analysts seeking to understand and predict market dynamics. Moreover, the transparency of linear regression allows stakeholders to interpret how each variable contributes to price changes, unlike black-box models whose internal decision processes may be opaque [3][4].

Despite its advantages, linear regression does have limitations, particularly when handling nonlinear, complex, or high-dimensional datasets often encountered in stock

**Corresponding Author:** M. Makmun Effendi, Universitas Pelita Bangsa, Indonesia (effendiyan@pelitabangsa.ac.id)
1. M. Makmun Effendi, Universitas Pelita Bangsa, Indonesia
2. Ahmad Turmuzi Zy, Universitas Pelita Bangsa, Indonesia
3. Isarianto, Universitas Pelita Bangsa, Indonesia

market data. Recent advances have therefore seen linear regression combined with other techniques such as moving averages to reduce noise, feature selection to improve model quality, and integration with machine learning methods to capture nonlinear relationships. These hybrid approaches extend the usability of linear regression, making it a versatile component of modern predictive analytics in finance [5][6]. Furthermore, the an increasing availability of alternative data sources like financial news, social media sentiment, and macroeconomic reports. It has opened new avenues for enhancing linear regression models. For instance, incorporating textual data through natural language processing (NLP) allows models to integrate market sentiment and qualitative factors that impact stock prices. This multidisciplinary approach represents an evolution from purely numerical modeling towards enriched, data-driven forecasting systems [1].

Several studies have also emphasized the importance of understanding inter-stock relationships and market trends beyond individual stock analysis. Multiple linear regression (MLR) enables the simultaneous modeling of various stocks and market indices, considering their interdependencies to produce more robust predictions. This is particularly important for portfolio management and diversification strategies where the interaction among assets affects overall risk and return profiles [7]. In summary, while machine learning and deep learning methods continue to grow in popularity, linear regression's fundamental role in stock price forecasting remains strong. Its combination of interpretability, low computational cost, and compatibility with other techniques ensures it will continue to be a valuable tool for financial analysts and researchers. The next section explores how these models have been applied and enhanced in recent research, highlighting the ongoing relevance and adaptability of linear regression in the evolving financial landscape [5], [8], [9], [1].

## 2. Related Works

A study proposed an MLR to construct a prediction model to effectively forecast short-term trends for NVDA, AMD, and INTC. After refining with correlation-based feature selection, two of the three models achieved statistically significant results. Specifically, the model's performance measured by $R^2$ was moderate for AMD (0.752) and strong for NVIDIA (0.837), while lower for Intel (0.596). These findings suggest that MLR when combined with relevant technical indicators, provides meaningful predictive insight into stock movements [5]. Another article applied multiple linear regression to forecast trends in major semiconductor stocks NVDA, AMD, and INTC. They also performed correlation analysis to evaluate inter-stock relationships, demonstrating that considering stock interdependencies can improve prediction accuracy. Their methodology emphasizes the importance of multi-asset modeling in financial forecasting and portfolio management [7].

An article extended this comparison by analyzing linear regression against several machine learning models including LightGBM, XGBoost, Random Forest, LSTM, and GRU for LONGi stock price prediction. Their results showed machine learning models generally outperform linear regression on complex datasets, but the latter's interpretability and efficiency remain valuable, especially for simpler or well-structured data [11]. From an Indonesian perspective, Alpianto et al. combined moving averages with linear regression to filter out noise and fluctuations in stock prices. Their method achieved low Mean Absolute Percentage Error (MAPE) and high correlation, demonstrating the effectiveness of hybrid approaches in practical investment contexts. Wilda et al. similarly applied simple linear regression to forecast PT Unilever Indonesia Tbk's stock price, attaining an accuracy level with MAPE of 2.65%, underscoring that even basic models can yield strong predictive performance when correctly applied [5], [12].

Other local studies applied multiple linear regression to analyze the effect of financial indicators on LQ-45 companies' stock prices, confirming the significant influence of PER, EPS, and ROE. Another work enhanced regression models by integrating K-Means clustering and moving average methods to address outliers, resulting in improved prediction accuracy and robustness. Similarly, Zapar et al. applied regression within a Knowledge Discovery in Databases (KDD) framework for Bank BCA stock price prediction, achieving consistent performance with low error rates [3], [6], [9]. Additional research explored macroeconomic influences on stock prices using multiple linear regression to study the impact of global economic variables such as crude oil prices, exchange rates, inflation, GDP growth, and composite stock indices on the JII70 Sharia stock index. Their findings emphasize the necessity of incorporating external economic factors in predictive models to capture broader market dynamics [4].

A paper utilized technical stock data including opening, highest, and lowest prices, as well as foreign net transactions in a multiple linear regression model for predicting Bank Rakyat Indonesia's stock price, showing that careful variable selection can significantly enhance forecast accuracy. Another work compared LR with ANN that shows ANN can obtain a better performance in complex datasets, and LR remains useful in less complicated scenarios [13][14]. Another paper evaluated LR, LSTM, and GRU for predicting Netflix stock prices with two decades of historical data. The experimental results displayed that DL models excelled at longer-term predictions, and LR remained effective for short-term forecasts. This comparison highlights LR's enduring relevance as a baseline and complementary model in complex predictive frameworks [10].

Another study combined data mining techniques with multiple linear regression to predict Netflix stock prices, obtaining strong evaluation metrics such as RMSE and MAE. Several studies have further confirmed the strengths and limitations of linear regression relative to machine learning approaches like RF, SVR, LASSO, KNN, and XGBoost, illustrating that linear regression remains a cornerstone method in the predictive modeling toolkit despite growing model complexity [15], [16], [17], [18], [19]. Those works illustrate the wide-ranging applications, adaptations, and ongoing relevance of linear regression in stock price prediction. They highlight how combining classical methods with modern data analytics and alternative data sources leads to better predictive performance and deeper market understanding [1]-[19].

# 3. Proposed Method

In this research, we propose the LR method for predicting the stock prices of KLBF as a statistical approach to measure the relationship between a dependent variable and independent variables. This method is particularly advantageous due to its simplicity and interpretability, making it accessible for users who may not have extensive expertise in machine learning modeling. By establishing a clear linear relationship, stakeholders can easily understand how changes in independent variables, such as historical prices and trading volumes, influence future stock prices.

The process begins with the identification of key independent variables that impact stock prices. In this study, we focus on historical stock prices, trading volumes, and other relevant financial indicators. The dependent variable is defined as the future stock price of KLBF. By analyzing historical data, we can derive a linear equation that best fits the observed data points, allowing us to make predictions about future stock prices based on established trends. This approach is particularly suitable for short-term predictions, where historical data can provide valuable insights into future price movements.

To ensure the robustness of the model, we will also implement data preprocessing techniques, including normalization and outlier detection, to enhance the quality of the input data. This preprocessing step is crucial as it helps to mitigate the impact of anomalies and ensures that the model is trained on high-quality data. By refining the dataset before applying the Linear Regression algorithm, we aim to improve the accuracy and reliability of the predictions generated by the model.

The general equation of a Linear Regression model is:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

Where:

1. $\hat{y}$ = predicted value (e.g., the future stock price of KLBF)
2. $\beta_0$ = intercept (constant term)
3. $\beta_1, \beta_2, \ldots, \beta_n$ = regression coefficients
4. $x_1, x_2, \ldots, x_n$ = independent variables (such as historical stock prices, trading volume, financial indicators, etc.)

The objective of the model is to minimize the Mean Squared Error (MSE), defined as:

$$MSE = (1/m) * \Sigma (y_i - \hat{y}_i)^2$$

Where:

1. m = number of observations
2. $y_i$ = actual value
3. $\hat{y}_i$ = predicted value

This formulation enables the model to find the best-fitting linear relationship between the input features and the target variable.

## 4. Experimental Setup

The experimental setup of this study includes data collection, preprocessing, model training, and evaluation. Initially, we collected historical stock price data for KLBF from Yahoo Finance, a reputable source for financial data. The dataset comprises 507 data points, covering the period from January 2020 to January 2022. This timeframe is selected to capture a range of market conditions, including periods of volatility and stability, which are essential for developing a robust predictive model.

Each data record consists of the following attributes: Date, Open, High, Low, Close, Adj Close, and Volume. For this research, we focused specifically on the “Low” price as the independent variable (X) and the “Close” price as the dependent variable (Y), based on their strong linear relationship observed in the preliminary correlation analysis. Once the data is collected, we proceed to the preprocessing phase, which involves cleaning the dataset to ensure its quality. This includes checking for missing values, inconsistencies, and noise within the data. Any missing values are addressed through appropriate imputation techniques, while irrelevant attributes are removed to focus on the most impactful factors influencing stock prices. The remaining attributes, such as date, open, high, low, and close prices, are retained for analysis. This careful selection of relevant features is critical for enhancing the model's predictive power.

We undergo data preprocessing in the following steps, this study conducts data import of CSV files containing raw stock data and filtering non-essential columns such as Adjusted Close and Volume to reduce noise and focus only on relevant predictors. The retained attributes were Date, Low, and Close. The dataset was examined for missing or null values and converted data Type such as the “Date” attribute was recognized as nominal and was excluded from the regression model since it does not contribute directly to prediction.

This study implements the Linear Regression model to train the model on the designated training subset and then predicts the "Close" price values on the testing subset. To assess its predictive accuracy, the evaluation uses RMSE, a widely accepted metric in regression analysis that quantifies the average magnitude of prediction error. RMSE provides a direct measure of how closely the predicted values align with the actual outcomes, making it a reliable indicator of model performance. Table 1 depicts the KLBF stock prices dataset with different categories.

Table 1: KLBF stock prices in several categories

| Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|
| 2020-01-02 | 1475 | 1500 | 1460 | 1480 | 1480 | 6,829,200 |
| 2020-01-03 | 1480 | 1480 | 1465 | 1470 | 1470 | 3,360,100 |
| 2020-01-06 | 1470 | 1480 | 1455 | 1460 | 1460 | 3,309,300 |
| 2020-01-07 | 1460 | 1475 | 1450 | 1465 | 1465 | 3,464,300 |

After preprocessing, we divided the dataset into training and testing subsets, with 90% of the data allocated for training the model and 10% reserved for testing its performance. This division is essential for maintaining the chronological order of the data, which is crucial for time series analysis. The model was trained on the training subset using RapidMiner's Linear Regression to generate predictions and evaluate the model using RMSE to measure the model's predictive accuracy.

# 5. Result and Analysis

In this study, we conducted a detailed analysis to identify key trends and fluctuations in stock prices. By comparing the predicted values with actual market data, we can observe that the model captures significant price movements, particularly during periods of market volatility. This ability to track price changes is crucial for investors seeking to make informed decisions based on short-term market dynamics. The analysis also highlights the importance of using historical data to inform future predictions, emphasizing the model's reliance on past trends to forecast future price movements.

Furthermore, a comparative analysis with other studies reveals that while more complex models, such as LSTM networks, may yield higher accuracy, the Linear Regression model provides a faster and more interpretable solution. The simplicity of the Linear Regression model allows for quick implementation and understanding, making it an attractive option for investors and analysts alike. This aspect is particularly important in contexts where time and resources are limited, as it enables stakeholders to make timely decisions based on the model's outputs. To further support the model's performance, the following table presents the accuracy assessment through RMSE across the test dataset. Table 2 depicts the actual and predicted closing prices, the error (difference), and the squared error for each prediction.

Table 2: Actual and predicted closing prices, the error, and the squared error

| No. | Date | Actual Close | Predicted Close | Error | Squared Error |
|---|---|---|---|---|---|
| 1 | 2021-11-18 | 1,615 | 1,632 | 17 | 289 |
| 2 | 2021-11-19 | 1,595 | 1,618 | 23 | 529 |
| 3 | 2021-11-22 | 1,600 | 1,622 | 22 | 484 |
| 4 | 2021-11-23 | 1,610 | 1,604 | -6 | 36 |
| 5 | 2021-11-24 | 1,610 | 1,627 | 17 | 289 |
| ... | ... | ... | ... | ... | ... |
| 51 | 2022-01-28 | 1,670 | 1,688 | 18 | 324 |
| | **Total** | | | | **27,105** |
| | **RMSE** | | | | **≈ 23.06** |

Table 2 presents a comparison between actual and predicted closing prices over a series of 51 trading days using a linear regression model. It includes the prediction error (difference between actual and predicted values) and the corresponding squared error for each entry. The total sum of squared errors is 27,105, resulting in a RMSE of approximately 23.06, indicating the model's average prediction deviation from actual values. This low error margin indicates a strong predictive performance and confirms the effectiveness of the Linear Regression approach to predict stock price prediction.

## 6. Conclusion

This research demonstrates that Linear Regression can successfully capture short-term price movements and provide valuable insights to navigate the complexities of the stock market. The evaluation results include the prediction error and corresponding squared error for each data point, yielding a total sum of squared errors of 27,105. This RMSE = 23.06 reflects a relatively low average deviation between predicted and actual closing prices. The low RMSE value demonstrates the model's strong predictive performance and supports its effectiveness for forecasting.

Moreover, the study highlights the importance of utilizing historical data in stock price prediction. By analyzing past trends and patterns, the Linear Regression model effectively identifies key factors that influence stock prices, enabling more informed investment decisions. While the model has certain limitations, such as its reliance on internal stock attributes and potential underperformance in non-linear scenarios, it serves as a strong baseline for stock price forecasting. This research underscores the relevance of Linear Regression in financial analysis, particularly in environments where data is constrained and interpretability is essential.

Finally, the comparative analysis with more complex models, such as LSTM reveals that while advanced techniques may offer higher accuracy, the Linear Regression model provides a faster and more interpretable solution. This aspect is particularly beneficial for investors and analysts who require timely insights without the need for extensive computational resources. Overall, the findings of this study contribute to the ongoing discourse on stock price prediction methodologies and emphasize the enduring significance in the field of financial analysis.

## Acknowledgment

We would like to express our sincere gratitude to all those who contributed to the successful completion of this research. First and foremost, we extend our appreciation to the data providers at Yahoo Finance for making historical stock data readily available, which was crucial for our analysis. Their comprehensive dataset allowed us to conduct a thorough investigation into the stock price prediction of KLBF.

We also wish to acknowledge the support and guidance from our academic advisors and peers, whose insights and feedback were instrumental in refining our methodology and improving the overall quality of the study. Their expertise in financial analysis and statistical modeling provided us with valuable perspectives that enhanced our research outcomes.

Finally, we recognize the importance of the financial community, whose interest in stock price prediction drives the relevance of this research. We hope that our findings will contribute positively to the field of financial analysis and assist investors in making informed decisions based on reliable predictive models. Thank you to everyone who played a role in this research endeavor.

# References


[1] H. A. K. Ihlayyel, N. M. Sharef, M. Z. A. Nazri, and A. A. bakar, “An Enhanced Feature Representation Based on Linear Regression Model for Stock Market Prediction,” *Intelligent Data Analysis*, vol. 22, no. 1, pp. 45–76, Jan. 2018, doi: 10.3233/IDA-163316.

[2] A. Hayes, “Multiple Linear Regression (MLR): Definition, Formula, and Example,” Investopedia. Accessed: Oct. 15, 2023. [Online]. Available: https://www.investopedia.com/terms/m/mlr.asp

[3] X. Jin and C. Yi, “The Comparison of Stock Price Prediction Based on Linear Regression Model and Machine Learning Scenarios,” in *Proceedings of the 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022)*, Atlantis Press, Dec. 2022, pp. 837–842. doi: 10.2991/978-94-6463-030-5_82.

[4] M. Rusu, A. Popescu, and D. Ionescu, “Performance Comparison of Various Regression Models in Predicting Apple Inc. Stock Prices,” *J Bus Res*, vol. 14, no. 1, pp. 112–125, 2024.

[5] Q. Wang, C. Xu, and T. Zhou, “Stock Price Prediction Based on Multiple Linear Regression,” *BCP Business & Management CMAM*, vol. 36, pp. 48–54, Jan. 2023, doi: https://doi.org/10.54691/bcpbm.v36i.3384.

[6] L. Xia, “Comparison of Linear Regression, RNN, and LSTM Models in Predicting Johnson & Johnson Stock Prices,” *Journal of Financial Engineering*, vol. 15, no. 2, pp. 89–102, 2023.

[7] Q. Wang, C. Xu, and T. Zhou, “Using Multiple Linear Regression Model to Predict the Trend of Stock Prices,” *BCP Business & Management*, vol. 38, pp. 142–149, 2023, doi: https://doi.org/10.54691/bcpbm.v38i.3384.

[8] Y. Zhou, “Stock Forecasting Based on Linear Regression Model and Nonlinear Machine Learning Regression Model,” in *Advances in Economics, Management and Political Sciences*, EWA Publishing, Jan. 2024, pp. 7–13. doi: 10.54254/2754-1169/57/20230364.

[9] S. Bhatta, R. Sharma, and P. Gupta, “Evaluating the Accuracy of Machine Learning Models in Predicting Stock Prices,” *International Journal of Machine Learning Applications*, vol. 8, no. 1, pp. 34–50, 2023.

[10] Y. Zhou, “Research on the Prediction of Stock Price Based on Regression and Deep Learning Models—Taking Netflix as an Example,” *Advanced Education, Management, and Political Science*, vol. 8119, pp. 1–6, 2024, Accessed: Oct. 16, 2023. [Online]. Available: https://www.ewadirect.com/proceedings/aemps/article/view/8119

[11] L. Jin and H. Yi, “Comparative Analysis of Prediction Effects of Linear Regression and Machine Learning Algorithms on Stock Price: a Case Study of LONGi,” in *Proceedings of the 2022 6th International Conference on Business and Information Management (ICBBEM)*, 2022, pp. 190–193. doi: https://doi.org/10.2991/978-94-6463-011-5_44.

[12] Z. Zhou, “Comparative Study of Linear Regression, LSTM, and GRU Models in Predicting Netflix Stock Prices,” *Journal of Computational Finance*, vol. 12, no. 3, pp. 201–215, 2024.

[13] J. Li, “Ordinary Least Squares (OLS), Ridge, and XGBoost Models for Predicting Stock Prices of Listed Companies,” *Journal of Financial Analysis*, vol. 10, no. 3, pp. 78–91, 2023.

[14] Y. Yin, “Comparison of KNN, Linear Regression, and LSTM Models in Predicting GM Stock Prices from 2013 to 2023,” *Journal of Financial Data Science*, vol. 7, no. 2, pp. 56–70, 2024.

[15] Investopedia, “Multiple Linear Regression (MLR) and its Application in Stock Price Prediction,” Investopedia. Accessed: Oct. 15, 2023. [Online]. Available: https://www.investopedia.com/terms/m/multiple-linear-regression.asp

[16] A. Bhatta, P. Poudyal, D. Kumar Maharja, and A. Thapa, “Assessing Machine Learning’s Accuracy in Stock Price Prediction,” *International Journal of Computer (IJC)*, vol. 49, no. 1, pp. 46–63, Sep. 2023, [Online]. Available: https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

[17] B. Xia, “Stock Price Predication Based on Linear Regression, RNN, LSTM,” *BCP Business & Management*, vol. 38, pp. 355–362, Mar. 2023, doi: 10.54691/bcpbm.v38i.3715.

[18] Y. Li, “Stock Price Prediction based on Multiple Regression Models,” in *2023 International Conference on Computer, Machine Learning, and Artificial Intelligence (CMLAI 2023)*, Highlights in Science, Engineering and Technology, Apr. 2023, pp. 657–662.

[19] Ștefan Rusu, M. I. Boloș, and M. Leordeanu, “COMPARATIVE ANALYSIS OF REGRESSION MODELS FOR STOCK PRICE PREDICTION: LINEAR, SUPPORT VECTOR, POLYNOMIAL, AND LASSO,” *Journal of Financial Studies*, vol. 9, no. 17, pp. 143–156, Nov. 2024, doi: 10.55654/JFS.2024.9.17.09.