ICEBIT
The 4th International Conference On Economics , Business and Information Technology
June 10, 2023 pp. 810-815

Stock Price Prediction of PT. Jasa Marga (Persero) Tbk Using Linear Regression Algorithm

Difari Afreyna Fauziah
Institute Technology and Science Mandala
Jl. Sumatra No. 118 - 120
Jember, Indonesia
085777583880
difa@stie-mandala.ac.id

Agung Muliawan
Institute Technology and Science Mandala
Jl. Sumatra No. 118 - 120
Jember, Indonesia
089610345572
agung@itsm.ac.id

Muhamat Abdul Rohim
Institute Technology and Science Mandala
Jl. Sumatra No. 118 - 120
Jember, Indonesia
085604946242
rohim@stie-mandala.ac.id

ABSTRACT

The Covid-19 pandemic that has occurred throughout the world has resulted in crises in various sectors, one of which is also affecting the economic sector. This has affected the movement of the index on several stocks listed on the Indonesian Stock Exchange (IDX). During the Covid-19 pandemic, stock price movements were very random with significant increases and decreases. Therefore, it is necessary to predict stock prices that can help investors to see investment prospects in the future. In this study, stock predictions from PT Jasa Marga (Persero) Tbk will be carried out using the Linear Regression algorithm. The linear regression method is usually used to make predictions by measuring the correlation between two or more variables. Historical data on the shares of PT Jasa Marga (Persero) Tbk used in this test were obtained from finance.yahoo.com from 2017-2023. The results of this study indicate that using the linear regression algorithm can predict prices with an RMSE value of 216.379.

Keywords : capital market, economy, forecasting, data mining

1. INTRODUCTION

Stocks are one of the most popular financial market instruments among the public. This instrument is the choice of investors because stocks can provide attractive profit levels. Economic development in Indonesia has slowed because of the covid-19 pandemic. This causes the movement of index movements in several stocks listed on the Indonesia Stock Exchange (IDX). The Indonesia Stock Exchange (IDX) is a trusted capital market institution in Indonesia. This institution is responsible for providing the infrastructure, regulations and mechanisms used for securities trading including stocks, bonds, and other financial instruments. IDX has a role in monitoring and regulating companies and regulating companies listed on the stock exchange.

PT Jasa Marga (Persero) Tbk is a public company engaged in the management of toll roads in Indonesia. PT Jasa Marga is one of the leading companies in the toll road infrastructure sector in Indonesia and has an extensive toll network in various regions. Jasa Marga has an important role in the construction, maintenance, and operation of toll roads. The company is involved in the construction of new toll roads, expansion of existing toll roads, maintenance of toll road infrastructure and management of services and supporting facilities for toll road users. As a public company, shares of PT Jasa Marga (Persero) Tbk are listed and traded on the Indonesia Stock Exchange with the stock code "JSMR".

The Covid-19 pandemic that occurred in Indonesia had a negative impact on various lines of business, including the freeway service sector which is managed by PT Jasa Marga (Persero) Tbk. Based on the company's records, in 2020 the company's revenue decreased by around 50% compared to normal conditions before Covid-19. One of the triggers for the decline in income was due to government policies related to large-scale social restrictions to reduce the spread of the Covid-19 virus. This caused a significant decline in the share price of JSMR shares as seen from the movement of the stock price index in 2020. However, after the Covid-19 pandemic subsided, JSMR's monthly toll revenues increased even higher than pre-pandemic revenues. This increase occurred because of positive sentiment on the Eid homecoming flow, which significantly boosted traffic volume.

Stock price movements are very random with increases and decreases, so stock price predictions are needed which can help investors to be able to see investment prospects in the future. The price of a stock is influenced by several factors, including the company's financial performance, industry conditions, the current economy, government policies, market sentiment and others. Investors often buy stocks with the hope that the value of these shares will increase over time, so that investors can benefit from price appreciation or distribution of dividends.

Stock investment has risks because the price of shares fluctuates quite a lot and investment can also provide profit or loss. Investors need to conduct research and analysis before investing in stocks and consider the objectives of the investment, risk tolerance, and the desired investment period. Therefore, in this study a stock price prediction was carried out which can help investors to see investment prospects. The results of these stock predictions can help investors to be able to make the right decisions. Therefore, the right algorithm is needed to produce the right stock price predictions. A previous study conducted by Himawan et al. (2022) explains that predicting stock prices using the linear regression method for the stock price of PT, Bank Rakyat Indonesia Tbk produces an RMSE value of

27.80, which means that this method is good enough to be used to predict the stock price. This research is in line with research conducted by Putra et al. (2022) which revealed that the application of the linear regression algorithm produced error values measured using MAPE on training data, test data and overall data of 13.751%, 13.773% and 13.775%. These three results indicate that the predictions from the model used by the researcher have a small error value and can be categorized as accurate. Based on previous research, researchers used a linear regression algorithm to predict stock prices at PT. Jasa Marga (Persero) Tbk.

Linear Regression is one of the algorithms used for predictions by using a straight line to describe the relationship between two or more variables (Ramadhan & Pamuji, 2022). The variables in this algorithm are divided into two, namely independent variables and dependent variable. This Linear Regression Algorithm describes the relationship between the dependent variable and one or more independent variables. The goal of this algorithm is to find the best regression line that can describe the linear relationship between these variables. The dataset used in this study has 7 attributes, namely Date, Open, High, Low, Close and Volume with a total of 1518 data. Based on this background, predictions will be made using the Linear Regression algorithm so that it is hoped that it can assist investors in making decisions on the Indonesian stock exchange.

2. METODOLOGI PENELITIAN

In this study, the research method used can be seen in Figure 1. In this method, data collection will become a dataset, which will then pass through the data preprocessing stage. The next stage is predicting stock prices using the Linear Regression algorithm. The prediction results from the Linear Regression algorithm are evaluated.

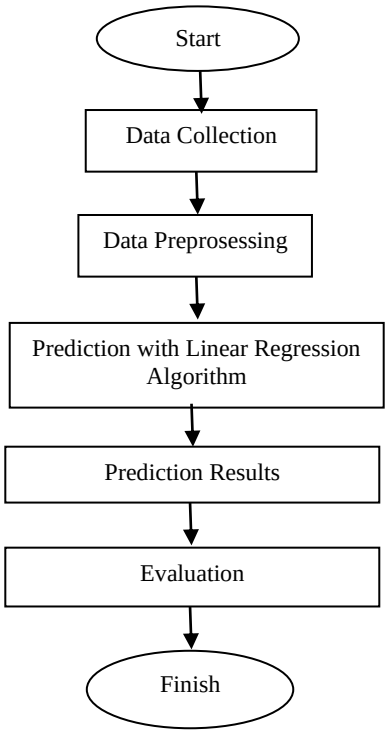

Figure 1 Method Research

2.1 Data Collection

In this study the data used was obtained from finance.yahoo.com. The data taken is PT Jasa Marga (Persero) Tbk data with a one-day time frame which is used as a research reference in analyzing. The amount of data taken is stock data with a range of 2 May 2017 to 24 May 2023 with a total of 1517 records. Based on the stock price dataset from PT Jasa Marga (Persero) Tbk, it consists of 4 attributes, namely date, open, high, low, volume. Meanwhile, the label is close. All these attributes apart from the label are things that affect the closing price of the stock. In this study using a comparison of training data as much as 70% and data testing as much as 30%. The details of the attributes contained in the data are shown in Table 1 as follows.

Tabel 1 Atribut Data Saham PT Jasa Marga (Persero) Tbk

No	Attribute	Data Type
1	Date	Date
2	Open	Interger

No	Attribute	Data Type
3	High	Interger
4	Low	Interger
5	Close	Interger
6	Adj Close	Interger
7	Volume	Interger

2.2 Data Preprosessing

Data preprocessing is one of the stages in data mining before entering the data processing stage (Reza Maulana & Devy Kumalasari, 2019). The data obtained in this study is raw data so it needs to be preprocessed first. There are several stages in data pre-processing, namely data cleaning, data integration, data transformation, and data reduction. The following stages of the data preprocessing process in this research can be seen in Table 2 and Table 3.

Table 2 Raw Data

Date	Open	High	Low	Close	Adj Close	Volume
02/05/2017	4.660.000.000	4.680.000.000	4.620.000.000	4.620.000.000	4.407.220.215	5928000
03/05/2017	4.640.000.000	4.640.000.000	4.590.000.000	4.610.000.000	4.397.680.176	5011700
04/05/2017	4.620.000.000	4.660.000.000	4.550.000.000	4.650.000.000	4.435.838.379	4865100
05/05/2017	4.660.000.000	4.750.000.000	4.600.000.000	4.740.000.000	4.521.693.359	7969500
08/05/2017	4.750.000.000	4.900.000.000	4.750.000.000	4.840.000.000	4.617.087.402	25406000
09/05/2017	4.840.000.000	4.860.000.000	4.750.000.000	4.780.000.000	4.559.851.074	7989000
10/05/2017	4.780.000.000	4.790.000.000	4.660.000.000	4.680.000.000	4.464.456.543	5709400
...	...	...	...	...	...	...
24/05/2023	3.460.000.000	3.540.000.000	3.460.000.000	3.530.000.000	3530000000	8494600

Source : https://finance.yahoo.com/

Table 3 Results of Preprocessing Data

Date	Open	High	Low	Close	Volume
02/05/2017	4660	4680	4620	4620	5928000
03/05/2017	4640	4640	4590	4610	5011700
04/05/2017	4620	4660	4550	4650	4865100
05/05/2017	4660	4750	4600	4740	7969500
08/05/2017	4750	4900	4750	4840	25406000
09/05/2017	4840	4860	4750	4780	7989000
10/05/2017	4780	4790	4660	4680	5709400
…	…	…	…	…	…
24/05/2023	3460	3540	3460	3530	8494600

Source: Data Processed 2023

2.3 Prediction Method with Linear Regression Algorithm

The forecasting process in this study uses the RapidMiner Studio 10.1 application. The RapidMiner application is a predictive analytics platform that is used to quickly mine information and data. The algorithm used in this study is the Linear Regression algorithm. Linear Regression is an algorithm used to measure the relationship between the correlations of two or more variables used for predictions through a straight line.

2.4 Evaluation

The results of data processing using the Linear Regression algorithm are in the form of Root Mean Square Error (RMSE). Root Mean Square Error is a measure that is usually used as the value of the difference between the value predicted by the model and the value obtained (Halimi & Kusuma, 2018). The individual difference between

the model label and the predicted results is called the residual. To calculate RMSE, that is by adding up all the squared residuals and then dividing by the amount of existing data and the results will be rooted.

3. RESULT AND DISCUSSION

3.1 Model Design

Modeling in this study was carried out using the RapidMiner Studio 10.1 application. The process begins by using the Read Excel operator, because the dataset in this study was collected in excel format. The next step is to select the excel dataset file to be used. There are several steps that must be passed, namely choosing the attributes to be used in the mining process, determining the data type format on the attributes and determining the attributes to be labeled. The next stage is using the set role operator, this operator is used to determine the date attribute in this study to be the id role. The next step is to use the windowing operator. Windowing is the process of forming a time series data structure into cross-sectional data. The window size used in this study is 5, which means that the last 5 days are processed to display forecasting results and display targets to be predicted (close attributes). After the model design is done, the next stage is split data which is used to divide the data into training data and data testing. In this study, the distribution of training data and testing data is 0.8 (training data) and 0.2 (testing data). The following is the design of the Linear Regression algorithm testing model in Figure 2.

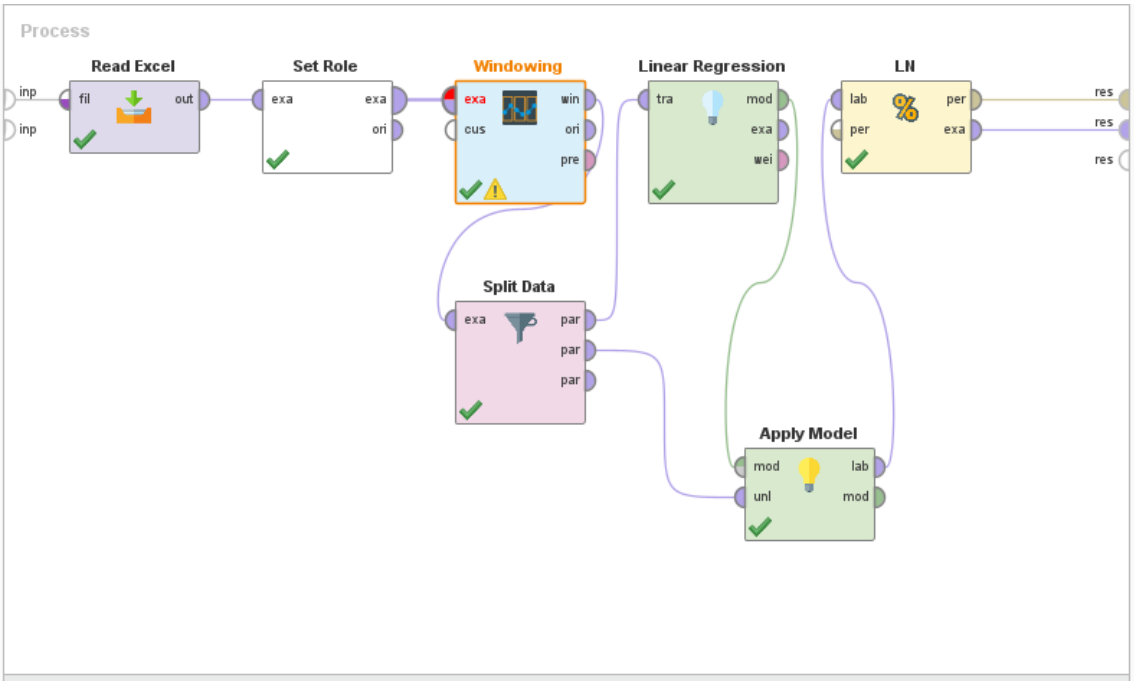

Figure 2 Model Desaign

3.2 Model Evaluation

The evaluation process carried out in this study uses the performance operator. The evaluation process is carried out using predetermined testing data as much as 30% of the dataset. The evaluation is carried out to determine the usability value that has been made with the previous steps, namely the data is processed using a linear regression operator followed by a performance operator to evaluate the performance of the model which provides a list of performance values automatically. The following is a graphical visualization of a comparison between the close attribute value and the close value from the predicted results using the Linear Regression algorithm.

Figure 3 Graph of Comparison of Close Values with Predicted Close Values

Open in Turbo Prep Auto Model Filter (454 / 454 examples): all

Row No.	Last Date in ...	Close + 1 (h...	prediction(C...	Open - 4	Open - 3	Open - 2	Open - 1	Op
441	Mar 8, 2023 1...	3250	3273.878	3250	3340	3340	3290	32
442	Mar 9, 2023 1...	3280	3384.629	3340	3340	3290	3250	32
443	Mar 14, 2023 ...	3060	3263.793	3250	3270	3300	3280	31
444	Mar 15, 2023 ...	3070	3295.516	3270	3300	3280	3180	31
445	Mar 16, 2023 ...	3150	3250.637	3300	3280	3180	3170	30
446	Mar 24, 2023 ...	3170	3057.385	3070	3100	3150	3160	31
447	Apr 3, 2023 1...	3290	3177.160	3190	3160	3230	3260	32
448	Apr 5, 2023 1...	3380	3230.971	3230	3260	3260	3280	32
449	Apr 10, 2023 ...	3390	3260.605	3260	3280	3290	3370	34
450	Apr 12, 2023 ...	3300	3307.849	3290	3370	3430	3410	33
451	May 4, 2023 1...	3320	3223.534	3240	3280	3310	3340	33
452	May 5, 2023 1...	3330	3248.588	3280	3310	3340	3320	33
453	May 11, 2023 ...	3470	3307.936	3350	3320	3330	3520	34
454	May 22, 2023 ...	3460	3375.644	3500	3370	3460	3600	33

ExampleSet (454 examples, 3 special attributes, 25 regular attributes)

Figure 4 Data Testing Prediction Results Using Linear Regression Algorithm

Figure 4 is the result of the prediction made by RapidMiner. The prediction column is the result of closing predictions generated using a linear regression algorithm. On May 22, 2023 the resulting price prediction for the closing price was 3375.644, but the actual closing price on that date was 3460. So the predicted closing price and the actual closing price had a difference of 84.356 rupiahs. This shows that the price difference between the two is not too far. The following results of the performance test of the linear regression model in this study are shown in Figure 5 which shows the calculation of the RMSE value of 216.379.

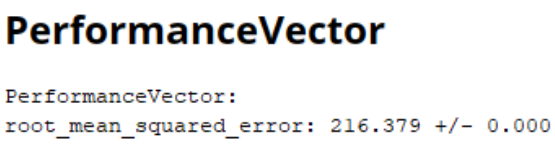

Figure 5 Performance Test Results

4. CONCLUSION

This research was conducted using the Linear Regression algorithm in making predictions related to the stock price of PT. Jasa Marga (Persero) Tbk using historical data obtained through the yahoo finance website in the period May 2017-May 2023. The linear regression algorithm can be used to predict stock prices at PT Jasa Marga (Persero) Tbk by using the ratio of data sharing between training data and testing data 80:20. The use of this algorithm produces an RMSE value of 216.379 using a dataset of 1517 records. With the stock price prediction model of PT. Jasa Marga (Persero) Tbk uses the Linear Regression algorithm so it can help investors to be able to see investment prospects in the future. For further research it is recommended to be able to utilize different algorithms and be able to use more and varied datasets.

5. REFERENCES

Halimi, I., & Kusuma, W. A. (2018). Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma Neural Network. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *4*(1), 24. https://doi.org/10.26418/jp.v4i1.25384

Himawan, I., Nurdiawan, O., & Dwilestari, G. (2022). *Algoritma Regresi Linier Berganda*. *10*(3).

Putra, J. S., Ramadhani, R. D., & Burhanuddin, A. (2022). Prediksi Harga Saham Bank Bri Menggunakan Algoritma Linear Regresion Sebagai Strategi Jual Beli Saham. *Journal of Dinda : Data Science, Information Technology, and Data Analytics*, *2*(1), 1–10. https://doi.org/10.20895/dinda.v2i1.273

Ramadhan, V. P., & Pamuji, F. Y. (2022). Analisis Perbandingan Algoritma Forecasting dalam Prediksi Harga Saham LQ45 PT Bank Mandiri Sekuritas (BMRI). *Jurnal Teknologi Dan Manajemen Informatika*, *8*(1), 39–45. https://doi.org/10.26905/jtmi.v8i1.6092

Reza Maulana, & Devy Kumalasari. (2019). Analisis Dan Perbandingan Algoritma Data Mining Dalam Prediksi Harga Saham Ggrm. *Jurnal Informatika Kaputama (JIK)*, *3*(1), 22–28. https://finance.yahoo.com/quote/GGRM.J