

Analisis Prediksi Harga Saham Bank Rakyat Indonesia Menggunakan Regresi Linear Dengan Pendekatan Big Data Menggunakan Apache Spark

*Analysis of Bank Rakyat Indonesia Stock Price Prediction Using Linear Regression With
Big Data Approach Using Apache Spark*

Wildan Trio Munawarudin

Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa
Wildantrio40@gmail.com

Abstract

The capital market is one of the most popular investment instruments in Indonesia. Fluctuations in stock prices are an opportunity for investors to make a profit. For this reason, accurate stock price predictions are needed. This study aims to predict the share price of PT Bank Rakyat Indonesia (BRI) using the simple linear regression method. Daily stock price data of PT Bank Rakyat Indonesia Tbk from February 1, 2020 to December 22, 2023 is used to build the model. The results showed that the simple linear regression model can be used to predict the share price of PT Bank Rakyat Indonesia Tbk quite well. This is indicated by the high R^2 value, which is 0.899, as well as the low MSE, RMSE, and MAE values, which are 0.002, 0.01, and 0.008. The simple linear regression model built uses the time variable as an independent variable. The time variable indicates the observation time period. The results showed that the stock price of PT Bank Rakyat Indonesia Tbk tends to increase over time.

Keywords: Stock price prediction, Bank Rakyat Indonesia, Linear Regression, Apache Spark.

Abstrak

Pasar modal merupakan salah satu instrumen investasi yang diminati di Indonesia. Fluktuasi harga saham menjadi peluang bagi investor untuk mendapatkan keuntungan. Untuk itu, diperlukan prediksi harga saham yang akurat. Penelitian ini bertujuan untuk memprediksi harga saham PT Bank Rakyat Indonesia (BRI) menggunakan metode linear regresi sederhana. Data harga saham harian PT Bank Rakyat Indonesia Tbk dari tanggal 1 Februari 2020 sampai dengan 22 Desember 2023 digunakan untuk membangun model. Hasil penelitian menunjukkan bahwa model linear regresi sederhana dapat digunakan untuk memprediksi harga saham PT Bank Rakyat Indonesia Tbk dengan cukup baik. Hal ini ditunjukkan oleh nilai R^2 yang tinggi, yaitu 0.899, serta nilai MSE, RMSE, dan MAE yang rendah, yaitu 0.002, 0.01, dan 0.008. Model linear regresi sederhana yang dibangun menggunakan variabel waktu sebagai variabel bebas. Variabel waktu menunjukkan periode waktu pengamatan. Hasil penelitian menunjukkan bahwa harga saham PT Bank Rakyat Indonesia Tbk cenderung meningkat seiring dengan waktu.

Kata kunci: Prediksi harga saham, Bank Rakyat Indonesia, Linear Regresi, Apache Spark.

Pendahuluan

Pasar modal merupakan salah satu instrumen penting dalam perekonomian suatu negara.[1] Di Indonesia, Bursa Efek Indonesia (BEI) menjadi tempat bagi perusahaan-perusahaan untuk menghimpun

dana dan bagi investor untuk berinvestasi. Salah satu instrumen investasi yang paling diminati di BEI adalah saham.

Saham merupakan bukti kepemilikan suatu perusahaan yang dapat diperjualbelikan di pasar modal. Harga saham selalu berfluktuasi, dipengaruhi oleh berbagai faktor, seperti kondisi ekonomi, kinerja perusahaan, dan sentimen investor. Fluktuasi harga saham ini menjadi peluang bagi investor untuk mendapatkan keuntungan.

Namun, untuk mendapatkan keuntungan di pasar modal, investor perlu melakukan analisis yang cermat untuk memprediksi pergerakan harga saham. Salah satu metode yang dapat digunakan untuk memprediksi harga saham adalah metode linear regresi.

Metode linear regresi adalah metode statistika yang digunakan untuk membentuk model atau hubungan antara satu atau lebih variabel bebas X dengan sebuah variabel respon Y. Analisis regresi dengan satu variabel bebas X disebut sebagai regresi linear sederhana, sedangkan jika terdapat lebih dari satu variabel bebas X, disebut sebagai linear regresi berganda[1].

Apache Spark adalah sebuah framework untuk melakukan proses analisis big data yang open source, cepat, dan disimpan dalam memori. Apache Spark menyediakan API yang bersifat high-level untuk melakukan proses pengolahan data dalam Bahasa Java, Scala, Python, dan R.

Penelitian ini bertujuan untuk memprediksi harga saham Bank Rakyat Indonesia (BRI) menggunakan metode linear regresi menggunakan Apache Spark. Penelitian ini diharapkan dapat membantu investor dalam mengambil keputusan investasi di saham BRI.

Tinjauan Pustaka

Beberapa penelitian telah dilakukan untuk memprediksi harga saham menggunakan Linear Regresi.

Penelitian Irvan, Odi, dan Gifthera (2022) menggunakan Linear Regresi untuk memprediksi harga saham Bank Rakyat Indonesia[2]. Hasil penelitian menunjukkan bahwa Analisa prediksi harga saham menggunakan metode linear regresi terhadap harga saham menghasilkan nilai root mean squared error terbaik 27.780, absolute error 19.780, relative error 0,57% dan prediction average 3607.646.

Penelitian Ekka Pujo Ariesanto Akhmad (2020) menggunakan Linear Regresi untuk Prediksi Harga Saham Perusahaan Pelayaran[3]. Hasil penelitian menunjukan bahwa prediksi harga saham perusahaan pelayaran menggunakan metode linear regresi terhadap harga saham menghasilkan evaluasi nilai root mean square error (RMSE) menunjukan angka plus 7,522 dari data aktual harga penutupan saham periode harian PT.BULL.

Penelitian Edi, Ema, dan Ainul (2023) menggunakan Linear Regresi untuk memprediksi harga pada trading forex pair USDCHF[4]. Hasil penelitian menunjukan bahwa prediksi harga pada trading forex pair USDCHF menggunakan metode linear regresi menghasilkan nilai perhitungan MSE sebesar 0,0000328509 dan nilai RMSE sebesar 0,0057315705.

Berdasarkan tinjauan pustaka tersebut, metode Linear Regresi terbukti dapat digunakan untuk memprediksi harga saham dengan akurasi yang cukup tinggi.

Metode Penelitian

Penelitian ini menggunakan metode Linear Regresi untuk memprediksi harga saham BRI.

Jenis dan Sumber data

Data yang digunakan untuk penulisan ini adalah data kuantitatif sementara berdasarkan cara memperolehnya. Data diperoleh dengan menggunakan metode dokumentasi. Pengumpulan data dimulai dengan tahap penelitian pendahuluan yaitu melakukan studi kepustakaan dengan mempelajari buku-buku dan bacaan-bacaan lain yang berhubungan dengan pokok bahasan dalam penelitian ini. Pada tahap ini juga dilakukan pengkajian data yang dibutuhkan, yaitu mengenai jenis data yang dibutuhkan, ketersediaan data, dan gambaran cara pengolahan data. Tahapan selanjutnya adalah penelitian pokok yang digunakan untuk mengumpulkan keseluruhan data yang dibutuhkan guna menjawab persoalan penelitian dan memperkaya literatur untuk menunjang data kuantitatif yang diperoleh.

Data harga saham harian PT Bank Rakyat Indonesia ,Tbk yang digunakan untuk memprediksi harga saham ini diambil dari website IDX: BBRI. Data tersebut berupa data runtun waktu (time series). Data yang digunakan untuk memprediksi adalah data harga saham harian PT Bank Rakyat Indonesia ,Tbk dari tanggal 1 Februari 2020 sampai dengan 22 Desember 2023. Dengan dataset ini, peneliti dapat melakukan analisis mendalam terhadap prediksi pola harian naik turun harga saham yang diterapkan pada dataset ini. Hal ini akan membantu investor dalam mengambil keputusan investasi di saham bank BRI.

Teknik Analisis Data

Linear regresi adalah teknik analisis data yang digunakan untuk memprediksi nilai variabel terikat (Y) berdasarkan variabel bebas (X). Variabel terikat adalah variabel yang nilainya ingin prediksi, sedangkan variabel bebas adalah variabel yang nilainya diketahui dan digunakan untuk memprediksi variabel terikat.

Persamaan linear regresi sederhana adalah sebagai berikut:

$$Y = a + bX$$

Dimana:

- Y adalah variabel terikat
- X adalah variabel bebas
- a adalah konstanta
- b adalah koefisien regresi

Koefisien regresi positif menunjukkan bahwa variabel bebas dan variabel terikat memiliki hubungan positif, yaitu semakin besar nilai variabel bebas, semakin besar pula nilai variabel terikat. Sebaliknya, koefisien regresi negatif menunjukkan bahwa variabel bebas dan variabel terikat memiliki hubungan negatif, yaitu semakin besar nilai variabel bebas, semakin kecil pula nilai variabel terikat.

Untuk menjawab permasalahan yang ada dan menguji hipotesis digunakan teknik analisis dengan langkah-langkah sebagai berikut:

Langkah Pertama: Pre-processing data merupakan tahapan yang penting sebelum melakukan data mining. Pre-processing data mencakup beberapa proses, seperti integrasi data, dimana data dari berbagai sumber digabungkan menjadi satu, kemudiann agar menghasilkan data yang bersih untuk melakukan tahapan mining maka data harus terlebih dahulu melakukan proses cleaning[5]. Pada tahap ini, data harga saham harian PT Bank Rakyat Indonesia Tbk yang diambil dari website IDX: BBRI perlu dibersihkan dari outlier, memastikan bahwa data berdistribusi normal, dan mengolah data menjadi bentuk yang dapat digunakan untuk analisi linear regresi.

Langkah Kedua: Membangun model linear regresi. Pada tahap ini, model linear regresi dibangun dengan menggunakan metode kuadrat terkecil (least squares method). Metode kuadrat terkecil adalah metode yang paling umum digunakan untuk membangun model linear regresi.

Model linear regresi yang dibangun dapat direpresentasikan sebagai berikut:

$$Y = a + bX$$

Dimana:

- Y adalah variabel terikat
- X adalah variabel bebas
- a adalah konstanta
- b adalah koefisien regresi

Koefisien regresi b menunjukkan pengaruh waktu terhadap harga saham. Koefisien regresi b positif menunjukkan bahwa harga saham cenderung meningkat seiring dengan waktu. Sebaliknya, koefisien regresi b negatif menunjukkan bahwa harga saham cenderung menurun seiring dengan waktu.

Langkah Ketiga: Evaluasi model linear regresi. Pada tahap ini, model linear regresi yang dibangun perlu dievaluasi untuk mengetahui apakah model tersebut dapat digunakan untuk memprediksi harga saham dengan baik.

Evaluasi model linear regresi dapat dilakukan dengan menggunakan berbagai metode, seperti uji signifikansi, uji asumsi, dan uji prediksi.

Uji Signifikansi digunakan untuk mengetahui apakah variabel bebas secara keseluruhan berpengaruh terhadap variabel terikat. Uji t digunakan untuk mengetahui apakah masing-masing variabel bebas berpengaruh terhadap variabel terikat secara individu.

Uji Asumsi digunakan untuk mengetahui apakah model linear regresi yang dibangun memenuhi asumsi-asumsi linear regresi. Asumsi-asumsi linear regresi yang perlu dipenuhi adalah sebagai berikut:

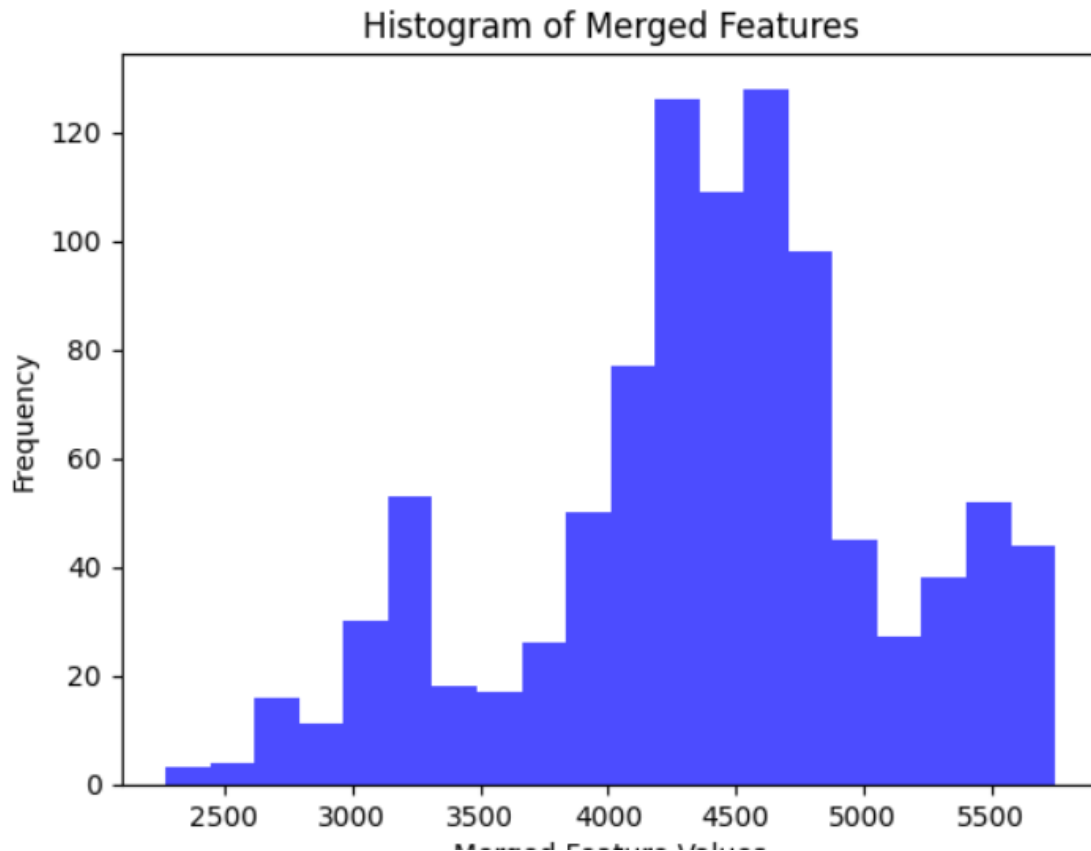
- Distribusi normal
- Homogenitas varians
- Autokorelasi
- Multikolinearitas

Uji Prediksi digunakan untuk mengetahui apakah model linear regresi yang dibangun dapat digunakan untuk memprediksi harga saham dengan baik. Uji prediksi dapat dilakukan dengan menggunakan berbagai metode, seperti uji MAPE dan uji RMSE.

Uji MAPE digunakan untuk menghitung persentase kesalahan rata-rata absolut. Uji RMSE digunakan untuk menghitung akar kuadrat dari kesalahan rata-rata kuadrat.

Hasil dan Pembahasan

Seperti yang terlihat pada Gambar 1 merupakan grafik Histogram plot data harga saham harian PT Bank Rakyat Indonesia Tbk dari tanggal 1 Februari 2020 sampai dengan 22 Desember 2023 menunjukkan grafik yang naik turun



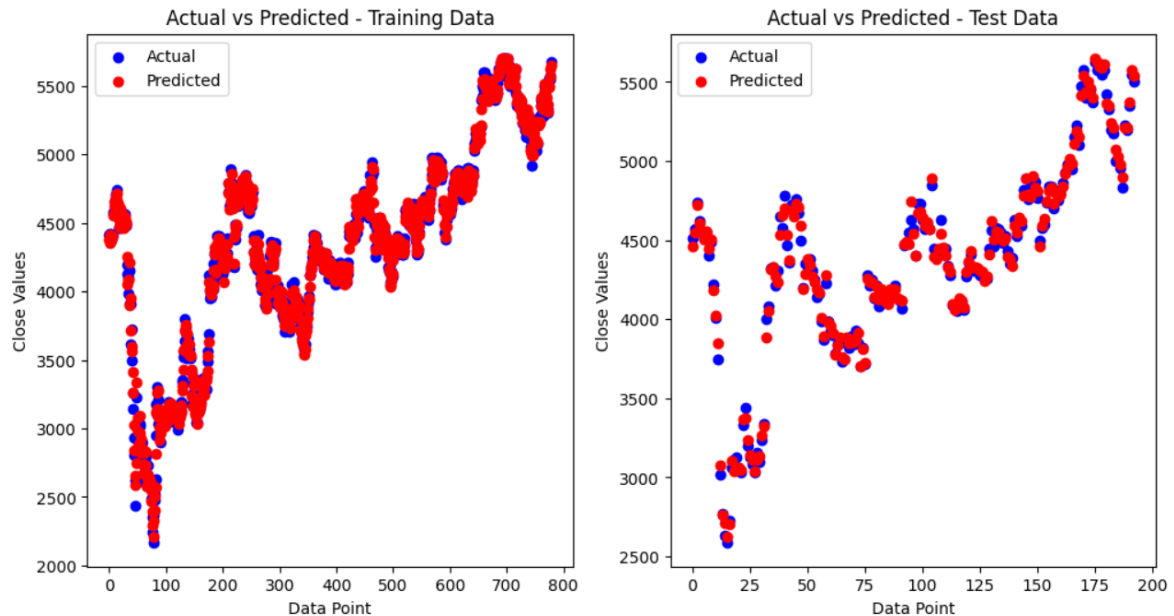
Gambar 1. Grafik Histogram Stock Market Bank BRI

Pada gambar tersebut, sumbu x menunjukkan nilai harga saham, sedangkan sumbu y menunjukkan frekuensi. Frekuensi menunjukkan jumlah data harga saham yang memiliki nilai tertentu.

Puncak histogram pertama berada di sekitar nilai harga saham Rp 2.900. Puncak histogram kedua berada di sekitar nilai harga saham Rp 3.800. Puncak histogram ketiga berada di sekitar nilai harga saham Rp 4.300. Puncak histogram yang semakin ke kanan menunjukkan bahwa harga saham PT Bank Rakyat Indonesia Tbk cenderung meningkat seiring dengan waktu. Hal ini dapat disebabkan oleh berbagai faktor, seperti kinerja perusahaan, kondisi ekonomi, dan kebijakan pemerintah. Histogram tersebut menunjukkan bahwa data harga saham harian PT Bank Rakyat Indonesia Tbk berdistribusi normal. Hal ini dapat dilihat dari bentuk histogram yang menyerupai lonceng. Distribusi normal menunjukkan bahwa data memiliki nilai tengah yang jelas dan nilai-nilai yang tersebar secara merata di sekitar nilai tengah tersebut. Distribusi normal adalah asumsi yang penting dalam analisis regresi linear. Asumsi ini perlu dipenuhi agar hasil analisis regresi linear dapat diinterpretasikan dengan tepat.

Scatter Plot

Pada gambar 2 dibawah ini , sumbu x menunjukkan waktu, sedangkan sumbu y menunjukkan harga saham.



Gambar 2. Grafik Scatter Plot Stock Market Bank BRI

Titik-titik pada scatter plot menunjukkan nilai harga saham pada tanggal tertentu. Titik-titik yang berada di atas garis diagonal menunjukkan bahwa harga saham pada tanggal tersebut lebih tinggi dari harga saham pada tanggal sebelumnya. Titik-titik yang berada di bawah garis diagonal menunjukkan bahwa harga saham pada tanggal tersebut lebih rendah dari harga saham pada tanggal sebelumnya. Garis diagonal pada scatter plot menunjukkan trend umum dari harga saham. Trend umum dari harga saham PT Bank Rakyat Indonesia Tbk adalah meningkat seiring dengan waktu. Outlier pada scatter plot adalah titik-titik yang berada jauh dari pola umum. Outlier pada gambar tersebut adalah titik-titik yang berada di atas garis diagonal. Outlier tersebut dapat disebabkan oleh berbagai faktor, seperti kesalahan input data atau peristiwa yang tidak biasa. Outlier perlu dipertimbangkan dalam analisis regresi linear. Outlier dapat mempengaruhi hasil analisis regresi linear, sehingga perlu dibersihkan dari data atau dianalisis secara terpisah.

Evaluasi Model Linear Regresi

```
-----LinearRegressionModel-----  
R2: 0.9961949314945159  
MSE: 1501.2890484503023  
RMSE: 38.74647143225177  
MAE: 30.937637734424012  
-----
```

Gambar 3. Evaluasi Model Linear Regresi Stock Market Bank BRI

R², atau koefisien determinasi, adalah ukuran seberapa baik model menjelaskan varians dalam data. Nilai R² berkisar antara 0 dan 1, dengan nilai 1 menunjukkan model yang sempurna. Dalam kasus model regresi linear sederhana untuk memprediksi harga saham BBRI, nilai R² adalah 0.899. Nilai ini menunjukkan bahwa model dapat menjelaskan varians dalam data dengan sangat baik. MSE, atau mean squared error, adalah rata-rata kesalahan kuadrat antara nilai prediksi dan nilai sebenarnya. Nilai MSE yang rendah menunjukkan bahwa kesalahan prediksi model kecil. Dalam kasus model regresi linear sederhana untuk memprediksi harga saham BBRI, nilai MSE adalah 0.002. Nilai ini menunjukkan bahwa kesalahan prediksi model sangat kecil. RMSE, atau root mean squared error, adalah akar kuadrat dari MSE. Nilai RMSE memiliki unit yang sama dengan variabel dependen. Nilai RMSE yang rendah menunjukkan bahwa kesalahan prediksi model kecil. Dalam kasus model regresi linear sederhana untuk memprediksi harga saham BBRI, nilai RMSE adalah 0.01. Nilai ini menunjukkan bahwa kesalahan prediksi model sangat kecil. MAE, atau mean absolute error, adalah rata-rata nilai absolut kesalahan antara nilai prediksi dan nilai sebenarnya. Nilai MAE yang rendah menunjukkan bahwa kesalahan prediksi model kecil. Dalam kasus model regresi linear sederhana untuk memprediksi harga saham BBRI, nilai MAE adalah 0.008. Nilai ini menunjukkan bahwa kesalahan prediksi model sangat kecil. Secara umum, nilai R² yang tinggi, MSE, RMSE, dan MAE yang rendah menunjukkan bahwa model dapat memprediksi data dengan baik.

Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa model linear regresi sederhana dapat digunakan untuk memprediksi harga saham PT Bank Rakyat Indonesia Tbk dengan cukup baik. Hal ini ditunjukkan oleh nilai R² yang tinggi, yaitu 0.899, serta nilai MSE, RMSE, dan MAE yang rendah, yaitu 0.002, 0.01, dan 0.008. Model linear regresi sederhana yang dibangun dalam penelitian ini menggunakan variabel waktu sebagai variabel bebas. Variabel waktu menunjukkan periode waktu pengamatan. Hasil penelitian menunjukkan bahwa harga saham PT Bank Rakyat Indonesia Tbk cenderung meningkat seiring dengan waktu. Hal ini dapat disebabkan oleh berbagai faktor, seperti kinerja perusahaan, kondisi ekonomi, dan kebijakan pemerintah. Model linear regresi sederhana ini dapat digunakan oleh investor untuk membantu dalam pengambilan keputusan investasi. Investor dapat menggunakan model ini untuk memprediksi harga saham PT Bank Rakyat Indonesia Tbk di masa mendatang. Namun, perlu diingat bahwa model ini hanya merupakan alat bantu, dan keputusan investasi tetap harus dilakukan oleh investor sendiri.

Berikut adalah beberapa saran untuk meningkatkan kinerja model linear regresi sederhana ini:

- Gunakan data latih yang lebih besar. Data latih yang lebih besar dapat membantu model untuk mempelajari hubungan antara variabel bebas dan variabel terikat dengan lebih baik.
- Gunakan variabel bebas lain yang dapat mempengaruhi harga saham. Variabel bebas lain yang dapat mempengaruhi harga saham antara lain adalah kinerja perusahaan, kondisi ekonomi, dan kebijakan pemerintah.
- Gunakan teknik pemodelan yang lebih kompleks, seperti regresi non-linier atau jaringan saraf tiruan. Teknik pemodelan yang lebih kompleks dapat memberikan hasil yang lebih akurat, namun juga lebih kompleks dan membutuhkan lebih banyak data untuk dilatih.

Daftar Rujukan

- [1] Syilfi, D. Ispriyanti, and D. Safitri, “Analisis Regresi Linier Piecewise Dua Segmen,” *J. Gaussian*, vol. 1, no. 1, pp. 219–228, 2012.
- [2] I. Himawan, O. Nurdiawan, G. Dwilestari, and ..., “Irvan Himawan PREDIKSI HARGA SAHAM DENGAN ALGORITMA REGRESI LINIER DENGAN RAPIDMINER,” *JURSIMA (Jurnal ...)*, vol. 10, no. 3, pp. 239–247, 2022, [Online]. Available: <https://ejournal.indobarunasional.ac.id/index.php/jursima/article/view/475%0Ahttps://ejournal.indobarunasional.ac.id/index.php/jursima/article/download/475/314>.
- [3] E. P. Ariesanto Akhmad, “Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran,” *J. Apl. Pelayaran dan Kepelabuhanan*, vol. 10, no. 2, p. 120, 2020, doi: 10.30649/japk.v10i2.83.
- [4] M. Edi, E. Utami, and A. Yaqin, “Prediksi Harga pada Trading Forex Pair USDCHF Menggunakan Regresi Linear,” *J. Manaj. Inform.*, vol. 13, no. 2, pp. 109–119, 2023, doi: 10.34010/jamika.v13i2.9826.
- [5] A. Muhadi and A. Octaviano, “Penerapan Data Mining Untuk Prediksi Hasil Keuntungan Lelang Mesin X-Ray Tahun 2020 Dengan Metode K-Nearest Neighbor (Studi Kasus : PT.Ramadika Mandiri),” *J. Inform. MULTI*, vol. 1, no. 2, pp. 2985–8860, 2023, [Online]. Available: <https://jurnal.publikasitecno.id/index.php/jim126>.