

Penerapan Algoritma Regresi Linear untuk Estimasi Harga Saham dalam Pengambilan Keputusan Investasi

Zaehol Fatah^{1,*}, Qurrotul A'yun²

¹Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

²Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

zaeholfatah@gmail.com, qurrotulayun1405@gmail.com

Article Info

Article history:

Received Nov 16, 2025

Accepted Nov 29, 2025

Published Jan 3, 2026

Kata Kunci:

Data Mining
Regresi Linear
Estimasi Harga Saham
RapidMiner
RMSE

ABSTRAK

Penelitian ini berfokus pada implementasi metode data mining dengan Regresi Linear (RL) untuk tujuan pemodelan dan estimasi harga saham menggunakan data historis. Pilihan terhadap algoritma RL didasarkan pada keunggulannya dalam struktur pemodelan linier, efisiensi komputasi, serta kemampuan menyajikan interpretasi hasil secara lugas. Dataset yang dimanfaatkan diperoleh dari platform Kaggle, yang berisi data harga saham harian dalam rentang waktu 2008–2015. Proses analisis data dilaksanakan menggunakan perangkat RapidMiner Studio, mencakup fase akuisisi data, pra-pemrosesan, implementasi algoritma, hingga penilaian kinerja model. Hasil estimasi dengan RapidMiner menunjukkan bahwa model RL berhasil memprediksi harga penutupan saham dengan tingkat akurasi yang tinggi, terlihat dari tingkat kesalahan yang minimal. Model tersebut telah teruji dengan mendapatkan nilai Root Mean Squared Error (RMSE) sebesar 1.229 dan Squared Error sebesar 1.511. Secara keseluruhan, dapat disimpulkan bahwa Regresi Linear adalah algoritma yang efektif untuk membangun model estimasi harga saham yang akurat, sehingga menjadi dasar yang kuat dalam pengambilan keputusan investasi.



Corresponding Author:

Qurrotul A'yun

1Sistem Informasi, Fakultas Sains dan Teknologi,

Universitas Ibrahimy, Indonesia

Email: *zaeholfatah@gmail.com

1. PENDAHULUAN

Kehidupan masyarakat modern kini sangat dipengaruhi secara signifikan oleh laju perkembangan yang pesat di sektor teknologi informasi, komunikasi, dan ekonomi. Saat ini, masyarakat memiliki berbagai alternatif untuk mengelola keuangannya, salah satunya melalui kegiatan investasi. Proses mengalokasikan sejumlah dana saat ini untuk mencapai hasil keuntungan finansial di masa depan dikenal sebagai investasi (Hidayati, Jurnal, and Islam 2017). Dari beragam bentuk investasi yang tersedia, investasi saham merupakan pilihan yang paling diminati. Saham sendiri didefinisikan sebagai instrumen finansial berupa surat berharga yang mengkonfirmasi kepemilikan modal sebagian dalam suatu entitas bisnis (Tania and Albertus Dwiyoga Widiatoro 2024). Salah satu instrumen di pasar modal yang menarik banyak peminat adalah saham, berkat tawaran potensi keuntungan yang signifikan. Secara umum, saham atau stock merupakan bukti kepemilikan individu terhadap perusahaan atau badan usaha tertentu. Nilai harga saham mencerminkan kinerja perusahaan sekaligus menjadi indikator kepercayaan investor terhadap prospek perusahaan tersebut. (Sari 2018) mendefinisikan harga saham sebagai total

dana yang harus dikeluarkan untuk membeli surat kepemilikan di sebuah entitas bisnis, berdasarkan pandangan Anoraga (2015:100). Nilai saham dapat berfluktuasi setiap waktu karena dipengaruhi oleh berbagai faktor, seperti kondisi fundamental perusahaan, kebijakan pemerintah, tingkat inflasi, suku bunga, serta situasi ekonomi global. Fluktuasi ini menyebabkan harga saham dapat berubah naik atau turun secara dinamis setiap harinya.

Menurut Tandelilin (2010:26), pasar modal adalah tempat yang mempertemukan orang-orang yang memiliki uang berlebih dengan orang-orang yang membutuhkan uang, melalui pembelian dan penjualan instrumen keuangan jangka panjang seperti obligasi dan saham (Ida Ayu Ika Mayuni, Gede Suarjaya 2018). Walaupun investasi saham menawarkan prospek imbal hasil yang menarik, risiko yang melekat padanya juga signifikan. Hal ini dikarenakan harga saham sangat rentan terhadap sentimen pasar dan berbagai kondisi ekonomi makro. Konsekuensinya, dibutuhkan metode analisis yang efektif untuk membantu investor mengidentifikasi pola pergerakan harga saham dan memperkirakan arah fluktuasinya. Dalam menganalisis pergerakan harga saham, data mining adalah salah satu metode yang bisa digunakan (Awaludin, Nuryadi, and Pribadi 2024). Untuk menemukan pola yang tersembunyi dan menjelajahi informasi penting dari sekumpulan data yang besar, diperlukan proses data mining yang menggunakan berbagai teknik seperti statistik, algoritma, dan kecerdasan buatan (Nur Khormarudin et al. 2022). Pendekatan ini memungkinkan analisis data keuangan dilakukan secara lebih sistematis untuk menemukan hubungan antar variabel yang berpengaruh terhadap harga saham (Ramadhani 2024). Berkat kemampuan ini, data mining berperan sebagai salah satu cara yang sangat efektif untuk membantu proses pengambilan keputusan yang didasarkan pada data di bidang investasi.

Seiring kemajuan teknologi analisis data, penerapan teknik data mining memberikan peluang besar bagi investor dan pelaku pasar modal untuk memperoleh informasi yang lebih akurat dalam menentukan strategi investasi (Awaludin and Mantik 2023). Melalui algoritma yang sesuai, data historis harga saham dapat diolah untuk membentuk model prediksi yang membantu memperkirakan arah tren pasar di masa mendatang. Dalam implementasi data mining, Regresi Linear merupakan salah satu metode yang umum diaplikasikan, baik dalam format sederhana maupun berganda (Awaludin 2020). Metode ini bertujuan untuk memformulasikan hubungan linier antara variabel terikat (harga penutupan saham) dengan beberapa variabel bebas. Di bidang analisis prediksi, Regresi Linear seringkali menjadi pilihan utama karena kemudahannya dalam implementasi, kecepatan komputasi, serta interpretasi hasil yang lugas (Khoiriyah and Fatah 2024). Penelitian terdahulu telah menunjukkan bahwa regresi linear mampu memberikan hasil estimasi yang akurat pada prediksi harga saham. Seperti penelitian oleh (Tania and Albertus Dwiyoga Widiatoro 2024) menerapkan algoritma ini dengan menggunakan perangkat lunak RapidMiner untuk menganalisis hubungan antar variabel harga saham Tesla Inc. (TSLA) dan menghasilkan tingkat akurasi yang sangat tinggi. Tingkat kesalahan prediksi terendah (RMSE = 56.699) dicapai ketika rasio data pelatihan dan pengujian ditetapkan 80%:20%. Ini adalah hasil yang diamati oleh Chang Hartono dan Dwiyoga Widiatoro (2024) dalam penelitian mengenai estimasi harga saham PT Unilever Indonesia Tbk, yang sekaligus memperkuat keandalan model regresi linear. Sementara itu, penelitian (Kuswanto et al. 2024) berjudul Estimasi Harga Saham dengan Algoritma Linear Regression menunjukkan bahwa penerapan regresi linear dalam kerangka data mining menggunakan RapidMiner menghasilkan model dengan nilai korelasi sebesar 0.884 dan R^2 sebesar 0.782. Temuan ini memperkuat bukti bahwa regresi linear efektif untuk membangun model estimasi numerik terhadap harga saham. Secara umum, hasil dari berbagai penelitian tersebut menunjukkan bahwa penerapan algoritma regresi linear pada proses data mining mampu menghasilkan model estimasi harga saham yang akurat dan efisien. Keterpaduan antara metode regresi linear dan pendekatan data mining memberikan dasar teoritis yang kuat untuk pengembangan sistem estimasi berbasis data historis (April et al. 2025).

Hubungan antara data mining dan regresi linear terletak pada peran regresi linear sebagai algoritma pemodelan dalam tahap analisis. Setelah data dibersihkan dan dipilih atribut yang relevan, regresi linear digunakan untuk membangun fungsi prediksi yang menggambarkan hubungan antara variabel input dan output (Sholeh, Nurnawati, and Lestari 2023). Karena regresi linear termasuk dalam kategori supervised learning yang berfokus pada prediksi nilai numerik, metode ini sangat sesuai untuk estimasi harga saham. Kombinasi kedua pendekatan ini memungkinkan terbentuknya model prediktif yang tidak hanya menggambarkan pola historis tetapi juga dapat divalidasi secara statistik untuk menghasilkan keputusan investasi yang lebih terukur (Usman et al. 2024). Dalam penelitian ini, metode regresi linear diimplementasikan menggunakan perangkat lunak RapidMiner. Pemilihan RapidMiner

didasarkan pada kemampuannya dalam mendukung analisis data mining secara visual, efisien, dan mudah digunakan tanpa memerlukan pemrograman tingkat lanjut (Citra Nursihah, Zaehol Fatah, and Rizki Hidayaturrochman 2025). RapidMiner juga mendukung berbagai metode analisis seperti klasifikasi, regresi, dan clustering, sehingga sangat cocok untuk penelitian yang berfokus pada estimasi harga saham. Berdasarkan pemaparan sebelumnya, penelitian ini berupaya untuk menghasilkan model estimasi harga saham yang akurat, efisien, dan informatif. Hal ini dicapai melalui aplikasi data mining dengan algoritma Regresi Linear, menggunakan RapidMiner sebagai alat bantu utama untuk memfasilitasi keputusan investasi.

2. METODE

Desain penelitian ini menggunakan pendekatan kuantitatif, yang secara spesifik menerapkan metode data mining untuk tujuan estimasi. Metode estimasi digunakan untuk memprediksi nilai numerik dari suatu variabel target berdasarkan variabel masukan yang relevan (Wibowo & Gunawan, 2023). Keunggulan pendekatan ini terletak pada kemampuannya untuk menyajikan temuan yang terukur dan netral melalui analisis data berbentuk numerik. Oleh karena itu, akurasi model prediksi dapat divalidasi dengan cara statistik (Sihombing & Ginting, 2022). Harga penutupan saham (closing price) ditetapkan sebagai variabel target yang akan diestimasi dalam studi ini. Pemilihan variabel-variabel dilakukan karena pengaruhnya yang dinilai signifikan terhadap fluktuasi harga saham, dan hal ini merupakan praktik umum dalam penelitian sejenis berbasis data mining (Marsandah & Widiatoro, 2024).

Algoritma Regresi Linear ditetapkan sebagai metode utama yang digunakan dalam fase estimasi. Secara matematis, kemampuan metode ini adalah memodelkan relasi linier antara satu atau lebih variabel bebas dan variabel terikat. Model yang dihasilkannya pun mudah untuk dipahami (Hartono & Widiatoro, 2024). Selain memiliki struktur sederhana, algoritma ini juga terbukti efektif pada berbagai studi prediksi harga saham karena mampu menghasilkan model yang akurat dan efisien ketika diimplementasikan menggunakan RapidMiner (Kuswanto et al., 2024). Untuk meninjau relevansi metode yang digunakan dalam penelitian ini, dilakukan analisis terhadap tiga studi terkini yang memanfaatkan algoritma Regresi Linear dalam konteks prediksi atau estimasi nilai numerik, khususnya harga saham. Ringkasan penelitian tersebut disajikan pada Tabel 1 sebagai state of the art, yang memuat masalah penelitian, metode yang digunakan, serta hasil utama yang dicapai masing-masing studi.

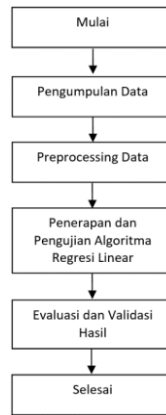
Tabel 1. Ringkasan State of the Art Penelitian Terkait

No	Nama Peneliti	Pembahasan	Hasil Penelitian
1	Tania & Albertus Dwiyoga Widiatoro (2024)	Meneliti masalah prediksi harga saham Tesla (TSLA). Metode yang digunakan adalah Regresi Linear dalam RapidMiner dengan pembagian dataset 80% data latih dan 20% data uji.	Regresi Linear memberikan akurasi prediksi yang baik dengan error yang stabil pada data saham TSLA, sehingga layak digunakan sebagai model estimasi harga saham.
2	Chang Hartono & Dwiyoga Widiatoro (2024)	Mengkaji pemodelan hubungan linier menggunakan algoritma Regresi Linear untuk memprediksi variabel numerik. Penelitian ini menekankan aspek interpretabilitas model dan relevansi variabel bebas dalam memengaruhi variabel terikat.	Penelitian menunjukkan bahwa Regresi Linear mampu memodelkan hubungan antarvariabel secara jelas, mudah dipahami, serta efektif untuk memprediksi nilai numerik dengan kompleksitas rendah.
3	Kuswanto et al. (2024)	Fokus pada estimasi harga saham menggunakan Linear Regression berbasis RapidMiner. Studi ini menilai hubungan antara variabel input dan harga saham melalui	Model menunjukkan performa tinggi dengan korelasi 0.884 dan $R^2 = 0.782$, membuktikan bahwa model linier efektif ketika pola hubungan

		pengujian statistik.	antarvariabel cenderung linier.
--	--	----------------------	---------------------------------

2.1 Tahapan Penelitian

Tahapan penelitian menggambarkan alur kerja sistematis dalam membangun model estimasi harga saham berbasis data mining. Setiap tahap dilakukan secara berurutan agar model yang dihasilkan valid, terukur, dan dapat direplikasi. Secara umum, alur penelitian ditunjukkan pada Gambar 1.



Gambar 1 Tahapan Penelitian

2.2 Pengumpulan Data

Akuisisi data adalah langkah esensial untuk memenuhi sasaran penelitian, yaitu membangun model estimasi harga saham menggunakan algoritma Regresi Linear. Data historis yang digunakan diperoleh dari platform Kaggle, yang mencakup pergerakan saham harian selama rentang waktu 2008–2015.

Pemilihan dataset dari Kaggle didasarkan pada kelengkapan, validitas, dan kemudahan akses, karena dataset tersebut banyak digunakan dalam penelitian berbasis data mining dan machine learning (Rochella, Lewa, and Witono 2024). Penggunaan data historis memungkinkan analisis pola fluktuasi harga dari waktu ke waktu sehingga menjadi dasar dalam proses estimasi harga penutupan saham.

Date	Open	High	Low	Close	Volume	Adj. Open	Adj. High	Adj. Low	Adj. Close	Adj. Volume
2004-08-19	100.01	104.06	95.96	100.33	44,659,000.00	50.16	52.19	48.13	50.32	44,659,000.00
2004-08-20	101.01	109.08	100.50	108.31	22,834,300.00	50.66	54.71	50.41	54.32	22,834,300.00
2004-08-23	110.76	113.48	109.05	109.40	18,256,100.00	55.55	56.92	54.69	54.87	18,256,100.00
2004-08-24	111.24	111.60	103.57	104.87	15,247,300.00	55.79	55.97	51.95	52.60	15,247,300.00
2004-08-25	104.76	108.00	103.88	106.00	9,188,600.00	52.54	54.17	52.10	53.16	9,188,600.00
2004-08-26	104.95	107.95	104.66	107.91	7,094,800.00	52.64	54.14	52.49	54.12	7,094,800.00
2004-08-27	108.10	108.62	105.69	106.15	6,211,700.00	54.22	54.48	53.01	53.24	6,211,700.00

Gambar 2. Data Set Keseluruhan Harga Saham

2.3 Preprocessing Data

Tahap pra-pemrosesan data dilakukan setelah proses pemilihan data (data selection) untuk memastikan dataset dalam kondisi bersih dan siap dianalisis lebih lanjut (Suriani 2023). Langkah ini penting karena kualitas data sangat memengaruhi keakuratan hasil analisis serta ketepatan model prediksi. Kegiatan yang tercakup dalam tahap pra-pemrosesan data dalam penelitian ini meliputi pembersihan, reduksi, dan normalisasi data. Langkah-langkah ini sangat krusial karena ditujukan untuk

memastikan konsistensi struktur data yang akan digunakan pada tahap pemodelan, sehingga hasil prediksi menjadi optimal (Patterns 2023).

2.4 Penerapan dan Pengujian Algoritma

Tahap penerapan dan pengujian merupakan inti dari penelitian, di mana model estimasi dibangun menggunakan algoritma Regresi Linear dengan bantuan perangkat lunak RapidMiner Studio. RapidMiner dipilih karena menyediakan antarmuka visual yang memudahkan pengguna dalam merancang alur analisis data mining secara sistematis tanpa menulis kode pemrograman.

Proses pemodelan diawali dengan melakukan impor dataset historis harga saham ke dalam perangkat lunak RapidMiner Studio. Setelah data dimuat, tahapan kunci adalah penentuan peran atribut; kolom harga penutupan (Close Price) ditetapkan sebagai Label (variabel target), sementara atribut lain (Volume, Open, High) ditetapkan sebagai prediktor. Proses pemisahan data dilakukan untuk membagi dataset menjadi dua bagian fungsional, yakni data pelatihan (training) dan data pengujian (testing) berdasarkan proporsi yang ditentukan, dilakukan dengan menambahkan operator Split Data. Menggunakan operator 'Linear Regression' untuk melatih model menggunakan data training (Awaludin, Yasin, and Risyda 2024). Data pelatihan kemudian dimasukkan ke operator Linear Regression untuk membangun fungsi model estimasi, yang selanjutnya diterapkan pada data pengujian melalui operator Apply Model. Evaluasi performa model dilakukan dengan operator Performance (Regression) untuk menghitung tingkat kesalahan estimasi (RMSE atau MAE) berdasarkan perbandingan antara nilai prediksi dan nilai aktual data uji. Dengan tahapan tersebut, model regresi linear yang dihasilkan dapat digunakan untuk memperkirakan nilai harga penutupan saham berdasarkan variabel prediktor yang telah dipilih (Rosika et al. 2024).

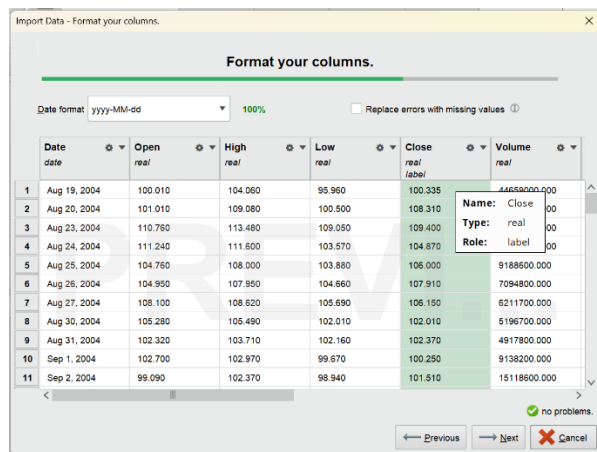
2.5 Evaluasi dan Validasi Hasil

Tahap terakhir adalah evaluasi dan validasi model guna menilai sejauh mana model regresi linear mampu melakukan estimasi harga saham dengan akurasi tinggi. Evaluasi dilakukan menggunakan operator Performance (Regression) di RapidMiner yang menghitung nilai kesalahan prediksi berdasarkan data uji.

Karena estimasi harga saham merupakan permasalahan regresi, penilaian difokuskan pada seberapa kecil perbedaan antara nilai aktual dan hasil prediksi. Nilai RMSE dan MAE digunakan sebagai indikator utama karena semakin kecil nilainya, semakin baik kinerja model dalam memprediksi harga saham (Sirojul Munir, A. 2024).

3. HASIL DAN PEMBAHASAN

Proses pengujian model estimasi harga saham dengan algoritma Regresi Linear dilakukan menggunakan perangkat lunak RapidMiner Studio. Langkah awal dilakukan dengan menambahkan operator Read CSV pada lembar kerja, kemudian memanfaatkan menu Import Configuration Wizard untuk memuat dataset harga saham yang diperoleh dari platform Kaggle. Setelah data berhasil dimasukkan, setiap atribut diatur sesuai dengan kebutuhan analisis agar dapat digunakan dalam proses pemodelan.



Gambar 3 Mengubah Atribut

Pada tahap ini, atribut Close ditetapkan sebagai Label karena berfungsi sebagai variabel target atau variabel dependen yang akan diprediksi. Sementara itu, atribut lain seperti Open, High, Low, dan Volume ditetapkan sebagai variabel bebas (independent variables) yang berperan sebagai prediktor. Setelah pengaturan atribut selesai dikonfigurasi, operator Split Data kemudian ditambahkan. Operator Split Data dimasukkan dengan maksud untuk memisahkan dataset menjadi dua kelompok penting: data pelatihan (training) dan data pengujian (testing).

Operator Split Data dikonfigurasi melalui menu Edit Parameter dengan proporsi 90% data pelatihan dan 10% data pengujian. Pembagian ini dimaksudkan agar model memiliki cukup data untuk mempelajari pola historis pergerakan harga saham, namun tetap menyisakan sebagian data untuk menguji kemampuan model dalam melakukan generalisasi.

Tahapan berikutnya adalah menambahkan operator Linear Regression guna membangun model estimasi. Operator tersebut dihubungkan dengan keluaran data pelatihan (tra) dari Split Data, sedangkan hasil model (mod) disambungkan ke operator berikutnya untuk diterapkan pada data uji. Proses ini menghasilkan model regresi linear yang menggambarkan hubungan antara variabel prediktor dengan harga penutupan saham.

LinearRegression

```

0.999 * Open
- 1.727 * Adj. Open
- 0.023 * Adj. High
+ 0.016 * Adj. Low
+ 1.735 * Adj. Close
- 0.000 * Adj. Volume
+ 0.281
                    
```

Gambar 4 Hasil Model Regresi Linear

Nilai koefisien regresi untuk setiap variabel input, sebagaimana disajikan pada Gambar 4, menjadi hasil utama dari pelatihan model. Analisis koefisien menunjukkan bahwa variabel Open dan Adjusted Close memiliki kontribusi positif yang paling dominan terhadap estimasi harga penutupan saham. Sebaliknya, Adjusted Open dan Adjusted High menampilkan hubungan yang negatif, sementara Adjusted Low memberikan kontribusi positif yang cenderung minor. Selain itu, karena nilai Adjusted Volume mendekati nol, dapat disimpulkan bahwa volume perdagangan tidak memiliki dampak signifikan pada pergerakan harga penutupan saham. Konstanta sebesar 0.281 ini merepresentasikan intercept yang terdapat pada model regresi linear yang dibangun.

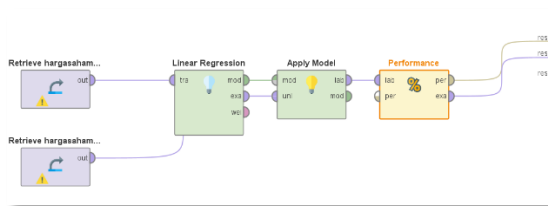
Setelah fase pelatihan, model tersebut diuji dengan data testing menggunakan operator Apply Model. Hasil dari penerapan model ini disajikan pada Gambar 5, di mana kolom Close berfungsi sebagai nilai aktual harga saham. Sementara itu, kolom prediction(Close) memperlihatkan hasil estimasi yang dihasilkan oleh model Regresi Linear.

Row No.	Close	prediction(C...	Date	Open	High	Low	Volume	Ex-Dividend
1	609.090	607.151	Feb 6, 2012	595.010	610.830	594.010	7351800	0
2	596.560	595.547	Nov 18, 2010	589	599.980	588.560	5174800	0
3	468.800	470.229	Jul 21, 2008	480.440	484.090	465.700	11791100	0
4	583.960	584.252	Jul 15, 2015	583.740	588.690	580.210	2073330	0
5	455.120	455.920	Apr 3, 2008	461.730	463.290	448.130	13543200	0

Gambar 5 Hasil dari Data Testing

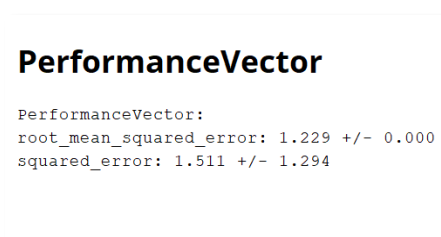
Dari hasil pengujian terlihat bahwa selisih antara nilai aktual dan nilai prediksi sangat kecil. Misalnya, nilai aktual Close sebesar 609,090 menghasilkan nilai prediksi 607,151, yang menunjukkan bahwa model memiliki kemampuan estimasi yang cukup baik. Hal ini mengindikasikan bahwa algoritma regresi linear mampu mengenali hubungan antara variabel bebas dan variabel target dengan efektif.

Untuk menilai performa model, digunakan operator Performance (Regression) sebagaimana ditunjukkan pada Gambar 6. Operator ini digunakan untuk menghitung tingkat kesalahan prediksi dan mengevaluasi akurasi model yang dihasilkan.



Gambar 6 Tahapan untuk Menguji Performa Regresi Linear

Setelah evaluasi dijalankan, RapidMiner menampilkan hasil metrik performa sebagaimana disajikan pada Gambar 7. Hasil dari proses tersebut mencatatkan nilai Root Mean Squared Error (RMSE) 1.229 ± 0.000 , diikuti dengan nilai Squared Error yang mencapai 1.511 ± 1.294 .



Gambar 7 Hasil Evaluasi Kinerja Regresi Linear

RMSE yang kecil mengindikasikan bahwa deviasi rata-rata antara nilai harga aktual dengan hasil estimasi adalah minimal. Sementara itu, rendahnya nilai Squared Error menunjukkan bahwa model memiliki stabilitas yang kuat dalam proses prediksinya. Temuan ini memvalidasi kapabilitas algoritma Regresi Linear untuk mengidentifikasi hubungan linier antar variabel secara efektif dan memberikan estimasi harga saham yang sangat akurat.

Rasio pembagian data 90% untuk pelatihan dan 10% untuk pengujian terbukti merupakan kombinasi terbaik. Proporsi ini mencegah terjadinya overfitting dan memungkinkan model untuk beradaptasi dengan baik pada data baru. Alokasi data pelatihan yang besar ini sangat penting karena memfasilitasi model dalam memahami pola historis saham secara lebih optimal.

Secara keseluruhan, penerapan algoritma Regresi Linear pada dataset harga saham menghasilkan model prediktif dengan tingkat kesalahan yang rendah serta hasil estimasi yang mendekati nilai aktual. Nilai RMSE sebesar 1.229 menjadi indikasi bahwa model memiliki tingkat akurasi tinggi, dan hasil prediksi yang konsisten menunjukkan bahwa model regresi linear layak digunakan dalam proses estimasi harga saham berbasis data historis..

4. KESIMPULAN

Penelitian ini dapat menyimpulkan bahwa data mining dengan algoritma Regresi Linear adalah pendekatan yang sangat efektif untuk membangun model estimasi harga saham, menghasilkan model prediksi yang akurat, stabil, dan interpretatif. Pemanfaatan RapidMiner Studio sangat membantu dalam proses analisis karena menyediakan alur pemodelan yang terstruktur mulai dari tahap pemrosesan data, pembagian dataset, pelatihan model, hingga evaluasi performa.

Model yang dikembangkan berhasil merepresentasikan hubungan linier antara variabel Open, High, Low, dan Volume terhadap harga penutupan saham (Close). Kinerja model terkonfirmasi dengan tingkat kesalahan prediksi yang minimal, angka-angka ini menunjukkan bahwa nilai Root Mean Squared Error (RMSE) adalah 1.229, dan nilai Squared Error tercatat sebesar 1.511. Dengan pembagian data sebesar 90% untuk pelatihan dan 10% untuk pengujian, model mampu mencapai keseimbangan yang baik antara kemampuan belajar dan performa uji tanpa indikasi overfitting.

Algoritma Regresi Linear pada proses estimasi harga saham yang berbasis data historis telah terbukti valid dan menunjukkan efisiensi yang tinggi. Model yang dihasilkan memiliki keunggulan signifikan dari segi efisiensi komputasi dan kemudahan dalam menginterpretasikan hasilnya. Untuk pengembangan riset selanjutnya, disarankan untuk melakukan perbandingan metode dengan algoritma machine learning lain, seperti Decision Tree atau Support Vector Regression (SVR). Perbandingan ini bertujuan untuk meningkatkan akurasi dan kapabilitas generalisasi model ketika dihadapkan pada dataset yang lebih kompleks.

DAFTAR PUSTAKA

- April, Volume Nomor, Siti Maisaroh, Ryci Rahmatil Fiska, Alamat Jl, Bathin Alam, Sungai Alam, and Bengkalis Riau. 2025. "Implementasi Algoritma Regresi Linier Berganda Untuk Prediksi Penjualan Di D' Kopikap Politeknik Negeri Bengkalis, Indonesia Dan Model Statistik Dalam Bisnis Kuliner Dapat Meningkatkan Akurasi Perencanaan Dan Efisiensi." 2019(April).
- Awaludin, Muryan. 2020. "Application Of Analytical Hierarchy Process Method For Employee Performance Evaluation At Pt Xyz." *JSI (Jurnal Sistem Informasi) Universitas Suryadarma* 7(1):137–50.
- Awaludin, Muryan, and Hari Mantik. 2023. "Penerapan Metode Servqual Pada Skala Likert Untuk Mendapatkan Kualitas Pelayanan Kepuasan Pelanggan." *Jurnal Sistem Informasi Universitas Suryadarma* 10(1).
- Awaludin, Muryan, Hepi Nuryadi, and Gerly Nasuha Pribadi. 2024. "Sistem Otomatisasi Laporan Untuk Optimalisasi Pelaporan Data Penelitian Dan Pengabdian Kepada Masyarakat Di Universitas Dirgantara Marsekal Suryadarma." 9675:1–7.
- Awaludin, Muryan, Verdi Yasin, and Fitria Risyda. 2024. "The Influence of Artificial Intelligence Technology, Infrastructure and Human Resource Competence on Internet Access Networks." *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi* 9(2):111–20. doi:10.25139/inform.v9i2.8109.
- Citra Nursihah, Zaehol Fatah, and Rizki Hidayaturrochman. 2025. "Penerapan Data Mining Untuk Penilaian Tes Hadrah Di Pesantren Salafiyah Syafi'iyah Menggunakan Metode Random Forest." *Jurnal Riset Teknik Komputer* 2(1):40–45. doi:10.69714/jfe1c445.
- Hidayati, Amalia Nuril, Malia. : Jurnal, and Ekonomi Islam. 2017. "Investasi: Analisis Dan Relevansinya Dengan Ekonomi Islam." *Jurnal Ekonomi Islam* 8(2):227–42.
- Ida Ayu Ika Mayuni. Gede Suarjaya. 2018. "SAHAM PADA SEKTOR MANUFAKTUR DI BEI." *E-Jurnal Manajemen Unud* 7(8):4063–93.
- Khoiriyah, Siti, and Zaehol Fatah. 2024. "Penerapan Algoritma Linear Regression Dalam Memprediksi Harga Rumah Menggunakan RapidMiner." *JUSIFOR : Jurnal Sistem Informasi Dan Informatika* 3(2):107–15. doi:10.70609/jusifor.v3i2.5684.
- Kuswanto, Andi Diah, Romanus Hinti Andro, Gibran Alif Januar, Rhyan Okto Sandrio, Nino Andera, and Titus Dwi Cahya. 2024. "Estimasi Harga Saham Menggunakan Algoritma Regresi Linear." *Jurnal Informatika* 11(1):1–8.
- Nur Khormarudin, Agus, Olivia Cantica Kevin Synagogue Panjaitan, Amin Septianingsih, Muhammad Faisal, Wiranti Sri Utami, Lisa Novia Ningsi, Heru Satria Tambunan, Charles Zai, and Teknik Komputer. 2022. "Teknik Data Mining: Algoritma K-Means Clustering." *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika* 1(2):116–23.
- Patterns, Historical. 2023. "Analisis Data Mining Untuk Prediksi Harga Saham : Perbandingan Metode Regresi Linier Dan Pola Historis." 4(2):278–88.
- Ramadhani, D. P. 2024. "Implementasi Data Mining Untuk Prediksi Pasar Saham: Studi Kasus." *Jurnal Dunia Data* 1(6):1–16.
- Rochella, Megan, Felicia Stevany Lewa, and Hans Witono. 2024. "Penerapan CRISP-DM Untuk Prediksi Harga Saham Starbucks Corporation Menggunakan Time Series Analysis." 4:214–26.
- Rosika, Herliana, Fakultas Teknik, Universitas Muhammadiyah Mataram, Fakultas Teknik, Universitas Mataram, Fakultas Rekayasa Sistem, Universitas Teknologi Sumbawa, Informasi Artikel, and Method K-means. 2024. "IMPLEMENTASI RAPIDMINER UNTUK CLESTERING DATA PENJUALAN PAKAIAN MENGGUNAKAN METODE K-MEANS." 5:221–31.
- Sari, Warsani Purnama. 2018. "Pengaruh Rasio Keuangan Terhadap Harga Saham Pada Perusahaan Manufaktur Go Public Yang Terdaftar Di Bursa Efek Indonesia." *Jurnal Ilmiah Skylandsea* 2(1):43–52.
- Sholeh, Muhammad, Erna Kumalasari Nurnawati, and Uning Lestari. 2023. "Penerapan Data Mining Dengan Metode Regresi Linear Untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner." *JISKA (Jurnal Informatika Sunan Kalijaga)* 8(1):10–21. doi:10.14421/jiska.2023.8.1.10-21.
- Sirojul Munir, A., et al. 2024. "Perbandingan Akurasi Algoritma Naive Bayes Dan Algoritma Decision Tree Dalam Pengklasifikasian Penyakit Kanker Payudara." 15(1):23–29.
- Suriani, Uci. 2023. "Penerapan Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma." 3(2):55–66.

- Tania, Sarah, and Albertus Dwiwoga Widianoro. 2024. "Implementasi Metode Regresi Linier Untuk Analisis Prediksi Harga Saham TSLA Di RapidMiner." *Sistematis : Jurnal Ilmiah Sistem Informasi* 1(1):27–35. doi:10.69533/67a7h013.
- Usman, Dandi Rakimad, Mukhlis Ramadhan, Masyuni Hutasuhut, Hendra Jaya, Rudi Gunawan, and Sri Kusnasari. 2024. "Implementasi Data Mining Untuk Memprediksi Pergerakan Harga Saham Bri Dengan Menggunakan Metode Regresi Linier Berganda." *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)* 7(1):151. doi:10.53513/jsk.v7i1.9605.