P-ISSN :2580-4316
E-ISSN :2654-8054 10.37817/ikraith-informatika.v9i2

PENGARUH DATA PREPROCESSING TERHADAP PERFORMA REGRESI LINIER DALAM PREDIKSI SAHAM

Aji Pangestu [1], Raden Teddy Iswahyudi [2]
[1,2] Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Esa Unggul
e-mail: mas7aji24@gmail.com, raden.teddy@esaunggul.ac.id

Abstrak

Penelitian ini membahas implementasi *data preprocessing* untuk meningkatkan performa model regresi linier dalam prediksi harga saham, dengan studi kasus pada saham 2012.TW periode Januari–Juni 2023. *Data preprocessing* menjadi tahap penting karena data finansial sering mengandung *missing values*, outlier, dan distribusi yang tidak seimbang. Tahapan preprocessing meliputi *data cleaning*, deteksi dan penanganan outlier, *feature engineering*, seleksi fitur, serta normalisasi. Model regresi linier kemudian dilatih dan diuji menggunakan *time series split* dengan evaluasi metrik R^2, MSE, dan MAPE. Hasil analisis menunjukkan bahwa regresi linier memiliki keterbatasan dalam menangkap dinamika harian ($R^2 = 0,39$), namun memberikan hasil yang lebih baik pada data mingguan ($R^2 = 0,62$) dan sangat kuat pada data bulanan ($R^2 = 0,95$). Nilai MSE yang relatif rendah pada ketiga skala data menunjukkan prediksi model cukup akurat terhadap tren harga. Dengan demikian, preprocessing berkontribusi signifikan terhadap peningkatan performa regresi linier, meskipun kompleksitas pasar saham menuntut pengembangan model yang lebih adaptif. Penelitian ini memberikan gambaran bahwa regresi linier dapat dijadikan baseline prediksi harga saham jangka pendek, serta membuka peluang integrasi dengan model pembelajaran mesin yang lebih canggih.

Kata Kunci: Data Preprocessing, Regresi Linier, Prediksi Saham, Evaluasi Model, Yahoo Finance

Abstract

This study discusses the implementation of data preprocessing to improve the performance of linear regression models in stock price prediction, with a case study on stock 2012.TW for the January–June 2023 period. Data preprocessing is essential since financial data often contain missing values, outliers, and imbalanced distributions. The preprocessing stages include data cleaning, outlier detection and handling, feature engineering, feature selection, and normalization. The linear regression model was then trained and tested using a time series split and evaluated with R^2, MSE, and MAPE metrics. The results show that linear regression has limitations in capturing daily dynamics ($R^2 = 0.39$), but performs better on weekly data ($R^2 = 0.62$) and is very strong on monthly data ($R^2 = 0.95$). The relatively low MSE across the three data scales indicates that the model's predictions are fairly accurate in capturing price trends. Thus, preprocessing contributes significantly to improving the performance of linear regression, although the complexity of the stock market requires the development of more adaptive models. This study illustrates that linear regression can serve as a baseline for short-term stock price prediction and opens opportunities for integration with more advanced machine learning models.

Keywords: Data Preprocessing, Linear Regression, Stock Prediction, Model Evaluation, Yahoo Finance

I PENDAHULUAN

Dalam analisis pasar saham, kualitas data memegang peran penting dalam menentukan akurasi model prediksi. Data keuangan sering kali bersifat kompleks, heterogen, dan mengandung ketidaklengkapan seperti *missing values*, outlier, maupun noise yang dapat mengganggu performa model prediktif. Oleh karena itu, proses *data*

preprocessing menjadi tahapan krusial untuk menghasilkan dataset yang lebih bersih, konsisten, dan representatif (Han et al., 2012). Regresi linier, meskipun merupakan metode statistik yang sederhana, masih banyak digunakan dalam pemodelan harga saham karena kemampuannya menjelaskan hubungan linier antara variabel independent, seperti volume perdagangan, indikator teknikal, maupun rasio keuangan dengan harga saham sebagai variabel dependen (Patel et al., 2015). Namun, tanpa tahapan preprocessing yang tepat, regresi linier dapat menghasilkan estimasi yang bias. Distribusi data yang tidak sesuai atau adanya multikolinearitas pada variabel independen dapat menurunkan validitas model (Gupta & Dhingra, 2019). Beberapa teknik preprocessing yang umum diterapkan meliputi pembersihan data (*data cleaning*), imputasi nilai hilang, deteksi dan penanganan outlier, serta normalisasi dan standardisasi fitur numerik. Normalisasi, misalnya, dapat meningkatkan kestabilan parameter dalam regresi linier dan meminimalkan pengaruh perbedaan skala antar variable (Jain et al., 2018). Selain itu, teknik seleksi fitur juga berperan penting dalam mengurangi dimensi data sekaligus meningkatkan interpretabilitas model (Tsai & Hsiao, 2010).

Dengan penerapan preprocessing yang tepat, regresi linier dapat mencapai performa prediksi yang lebih baik. Beberapa penelitian menunjukkan bahwa pengolahan data awal yang optimal mampu meningkatkan nilai R^2, menurunkan error metrik seperti *Mean Squared Error (MSE)*, serta memperkuat reliabilitas model dalam konteks prediksi pasar saham (Patel et al., 2015). Oleh karena itu, *data preprocessing* tidak hanya berfungsi sebagai langkah persiapan teknis, melainkan juga sebagai fondasi metodologis untuk memastikan akurasi dan stabilitas hasil prediksi.

II Studi Kajian Literasi

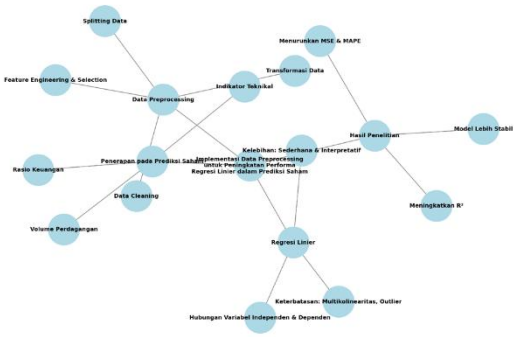
Kajian literatur mengenai implementasi *data preprocessing* dalam meningkatkan performa model regresi linier pada prediksi harga saham menunjukkan bahwa keberhasilan model sangat dipengaruhi oleh kualitas data yang digunakan. *Data preprocessing* merupakan tahap fundamental dalam proses penambangan data karena mampu mengatasi masalah *missing values*, inkonsistensi, dan noise yang umum ditemukan dalam data finansial (Han et al., 2012). Tanpa tahapan ini, model prediksi rentan menghasilkan hasil yang bias dan kurang akurat. Dengan menyoroti pentingnya normalisasi data dalam meningkatkan stabilitas parameter pada algoritma regresi linier (Jain et al., 2018). Hal ini sejalan dengan temuan hasil kajian (Patel et al., 2015) yang membuktikan bahwa teknik persiapan data, termasuk normalisasi dan pembuangan *outlier*, dapat secara signifikan memperbaiki kinerja prediksi harga saham serta menurunkan error metrik seperti MSE. Temuan tersebut menegaskan bahwa pengolahan awal data bukan sekadar langkah teknis, melainkan fondasi penting dalam mengoptimalkan performa model.

Selain itu, kajian (Gupta & Dhingra, 2019) menekankan pada *feature selection* sebagai strategi efektif untuk mengurangi multikolinearitas dalam regresi linier. Dengan memilih variabel prediktor yang relevan, model menjadi lebih sederhana, interpretatif, dan stabil dalam menghasilkan estimasi harga saham. Selanjutnya, menunjukkan bahwa penggabungan beberapa metode seleksi fitur dapat meningkatkan akurasi prediksi sekaligus mengurangi kompleksitas model (Tsai & Hsiao, 2010), sehingga regresi linier dapat bersaing dengan model yang lebih kompleks dalam konteks data keuangan. Dari keseluruhan literatur, dapat disimpulkan bahwa preprocessing data bukan hanya berfungsi untuk *cleaning* dataset (Agustina et al., 2025), tetapi juga untuk meningkatkan reliabilitas, interpretabilitas, dan performa regresi linier. Penerapan preprocessing yang tepat memberikan kontribusi signifikan pada peningkatan akurasi prediksi harga saham, serta menjadikan regresi linier tetap relevan di tengah berkembangnya model-model pembelajaran mesin yang lebih kompleks. Mind map ruang lingkup (gambar 1) menggambarkan hubungan antara tahapan preprocessing, konsep regresi linier, penerapannya pada prediksi saham, serta hasil penelitian yang relevan.

III Metodologi Penelitian

Metodologi penelitian ini dirancang untuk menganalisis pengaruh implementasi *data preprocessing* (Mirfan et al., 2024) terhadap peningkatan performa regresi linier dalam prediksi harga saham. Metode yang digunakan bersifat kuantitatif (gambar 3) dengan

pendekatan *experimental design*, di mana data historis harga saham dan variabel terkait diproses melalui tahapan preprocessing sebelum dilakukan pemodelan regresi linier (Adhy et al., 2025; Agustina et al., 2025).



Gambar 1. Pemetaan Referensi Hasil Studi Kajian Literasi

3.1 Pengumpulan Data

Data diperoleh dari catatan historis pasar saham yang meliputi harga penutupan, volume perdagangan, dan rasio keuangan perusahaan. Penggunaan data sekunder ini sejalan dengan praktik umum dalam penelitian prediksi saham (Patel et al., 2015), yang memanfaatkan data pasar untuk menguji efektivitas teknik persiapan data terhadap kinerja model prediksi, dengan mengambil sample kode saham 2012.TW, periode Januari – Juni 2023.

3.2 Data Preprocessing

Tahap ini melibatkan beberapa langkah penting:

a) Data Cleaning: Mengatasi *missing values* dengan metode imputasi (mean/median/ interpolasi) serta menghapus atau mengoreksi anomali data (Han et al., 2012).
b) Outlier Detection & Handling: Identifikasi data ekstrem yang dapat memengaruhi koefisien regresi.
c) Feature Engineering: Membuat variabel turunan seperti *moving average*, volatilitas, atau *rate of return* untuk memperkaya informasi prediktif.
d) Feature Selection: Pemilihan variabel relevan menggunakan uji korelasi atau *Variance Inflation Factor (VIF)* untuk mengurangi multikolinearitas (Gupta & Dhingra, 2019; Tsai & Hsiao, 2010).
e) Transformasi Data: Normalisasi atau standardisasi untuk menyeragamkan skala variabel numerik, yang terbukti dapat meningkatkan stabilitas parameter regresi (Jain et al., 2018).

3.3 Pemodelan Regresi Linier

Model regresi linier (Fairuzabadi et al., 2024) dibangun dengan harga saham sebagai variabel dependen, sementara variabel independen berupa indikator pasar dan keuangan yang telah melalui preprocessing. Model dilatih menggunakan *training set* dan diuji pada *testing set* dengan metode *time series split* untuk menghindari kebocoran data masa depan. Bentuk Umum Regresi Linier

$$\hat{y} = \beta_0 + \beta_1 x$$

dimana:

- $\hat{y}$= prediksi harga penutupan (Close)
- x= hari ke-n (misal: 0 untuk 8 Juni, 1 untuk 9 Juni, dst.)
- β_0= intercept (nilai saat $x = 0$)
- β_1= slope/kemiringan garis (kenaikan harga per hari)

Gambar 2. Data Statistik (Yahoo Finance, 2023)

3.4 Evaluasi Model

Kinerja model dievaluasi menggunakan metrik statistik, antara lain:

a) R^2 dan *Adjusted* R^2 untuk mengukur proporsi variasi yang dapat dijelaskan model.
b) MSE dan *Mean Absolute Percentage Error* (MAPE) untuk mengukur kesalahan prediksi.

Sebagaimana kombinasi metrik ini memberikan gambaran menyeluruh mengenai akurasi dan stabilitas model (Patel et al., 2015).

3.5 Arsitektur Dan Implementasi Model

P-ISSN :2580-4316
E-ISSN :2654-8054

Model prediksi yang digunakan dalam penelitian ini adalah regresi linier, dipilih karena mampu memodelkan hubungan linear antara variabel independen (fitur) dengan variabel dependen (target) secara sederhana dan mudah diinterpretasikan. Implementasi dilakukan menggunakan **libray Scikit-learn** dengan tahapan sebagai berikut:

Menentukan Fitur dan Target: variabel independen (X) terdiri atas harga pembukaan, harga tertinggi, harga terendah, harga penutupan yang disesuaikan, serta volume perdagangan; sedangkan variabel dependen (Y) adalah harga penutupan saham.

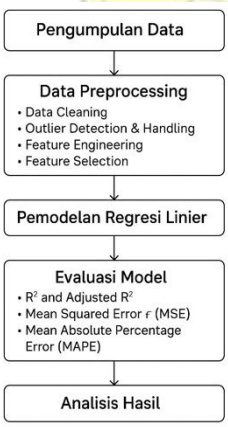
Pembagian Data: dataset dibagi menjadi data latih dan data uji dengan perbandingan 80% untuk pelatihan dan 20% untuk pengujian, guna mengevaluasi kemampuan generalisasi model.

Pelatihan Model: model regresi linier dilatih menggunakan data latih untuk mempelajari hubungan antara fitur dan target.

Evaluasi Model: kinerja model diuji pada data uji dengan menggunakan *metrik* MSE dan Koefisien Determinasi (R^2) untuk menilai akurasi prediksi.

3.6 Analisis Hasil

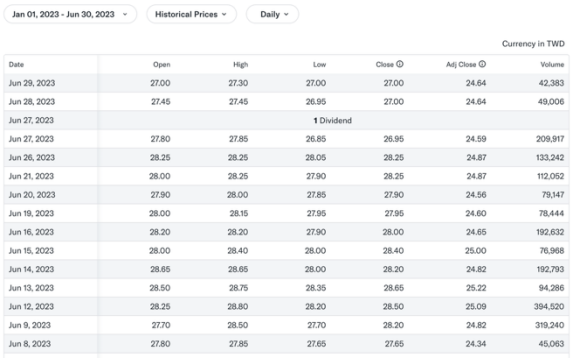
Hasil evaluasi dibandingkan antara model regresi linier dengan dan tanpa preprocessing. Analisis ini bertujuan untuk mengetahui sejauh mana preprocessing berkontribusi terhadap peningkatan performa prediksi harga saham.



Gambar 3. Metodologi

IV Pengumpulan, Hasil dan Pembahasan

4.1 Data Harian



Jan 01, 2023 - Jun 30, 2023 | Historical Prices | Daily

Currency in TWD

Date	Open	High	Low	Close	Adj Close	Volume
Jun 29, 2023	27.00	27.30	27.00	27.00	24.64	42,383
Jun 28, 2023	27.45	27.45	26.95	27.00	24.64	49,006
Jun 27, 2023	1 Dividend					
Jun 27, 2023	27.80	27.85	26.85	26.95	24.59	209,917
Jun 26, 2023	28.25	28.25	28.05	28.25	24.87	133,242
Jun 21, 2023	28.00	28.25	27.90	28.25	24.87	112,052
Jun 20, 2023	27.90	28.00	27.85	27.90	24.56	79,147
Jun 19, 2023	28.00	28.15	27.95	27.95	24.60	78,444
Jun 16, 2023	28.20	28.20	27.90	28.00	24.65	192,632
Jun 15, 2023	28.00	28.40	28.00	28.40	25.00	76,968
Jun 14, 2023	28.65	28.65	28.00	28.20	24.82	192,793
Jun 13, 2023	28.50	28.75	28.35	28.65	25.22	94,286
Jun 12, 2023	28.25	28.80	28.20	28.50	25.09	394,520
Jun 9, 2023	27.70	28.50	27.70	28.20	24.82	319,240
Jun 8, 2023	27.80	27.85	27.65	27.65	24.34	45,063

Gambar 4. Data Per Hari (Yahoo Finance, 2023)

MSE = 0.2767

Nilai ini menunjukkan bahwa rata-rata kesalahan kuadrat antara nilai aktual dan prediksi model cukup kecil. Dalam konteks harga saham yang berada dalam kisaran sekitar 27–28 TWD, nilai MSE ini tergolong rendah, yang berarti prediksi model cukup dekat dengan nilai aktual secara numerik.

R^2 = 0.3902

39% dari variasi harga saham dapat dijelaskan oleh model linier berdasarkan urutan hari. Sisanya, sekitar 61% variasi harga tidak bisa dijelaskan oleh model ini. Hal ini menunjukkan bahwa model linier kurang cocok untuk menggambarkan dinamika harga harian saham, yang umumnya bersifat non-linier dan dipengaruhi oleh berbagai faktor lain seperti berita pasar, volume transaksi, sentimen investor, dan kondisi makroekonomi.

Model Regresi Linier

$$\hat{y} = 28.2875 - 0.0932\,x$$

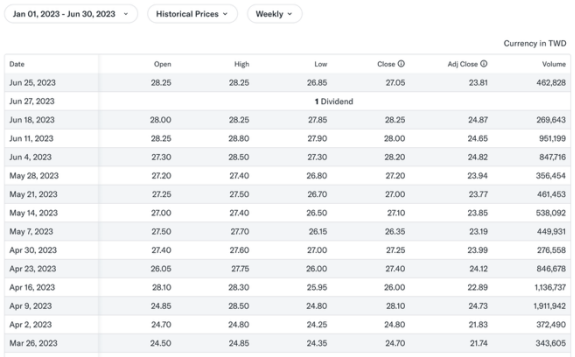
Interpretasi;

Intercept (β_0) = 28.2875, perkiraan harga penutupan pada hari ke-0 (yaitu 8 Juni 2023).

Slope (β_1) = - 0.0932, rata-rata harga saham turun sekitar 0.0932 TWD per hari selama periode ini menunjukkan tren penurunan ringan.

4.2 Data Mingguan dan Evaluasi Model

P-ISSN :2580-4316
E-ISSN :2654-8054
10.37817/ikraith-informatika.v9i2



Jan 01, 2023 - Jun 30, 2023 | Historical Prices | Weekly

Currency in TWD

Date	Open	High	Low	Close	Adj Close	Volume
Jun 25, 2023	28.25	28.25	26.85	27.05	23.81	462,828
Jun 27, 2023			1 Dividend			
Jun 18, 2023	28.00	28.25	27.85	28.25	24.87	269,643
Jun 11, 2023	28.25	28.80	27.90	28.00	24.65	951,199
Jun 4, 2023	27.30	28.50	27.30	28.20	24.82	847,716
May 28, 2023	27.20	27.40	26.80	27.20	23.94	356,454
May 21, 2023	27.25	27.50	26.70	27.00	23.77	461,453
May 14, 2023	27.00	27.40	26.50	27.10	23.85	538,092
May 7, 2023	27.50	27.70	26.15	26.35	23.19	449,931
Apr 30, 2023	27.40	27.60	27.00	27.25	23.99	276,558
Apr 23, 2023	26.05	27.75	26.00	27.40	24.12	846,678
Apr 16, 2023	28.10	28.30	25.95	26.00	22.89	1,136,737
Apr 9, 2023	24.85	28.50	24.80	28.10	24.73	1,911,942
Apr 2, 2023	24.70	24.80	24.25	24.80	21.83	372,490
Mar 26, 2023	24.50	24.85	24.35	24.70	21.74	343,605

Gambar 5. Data Per Minggu (Yahoo Finance, 2023)

MSE = 0.6894

Nilai ini menunjukkan bahwa rata-rata kuadrat kesalahan prediksi dari model adalah sekitar 0.69 TWD², yang relatif kecil jika dibandingkan dengan harga penutupan dalam rentang 24.7 – 28.25 TWD. Ini berarti bahwa secara numerik, prediksi model cukup mendekati nilai aktual.

$R^2 = 0.6212$

62.12% variasi harga saham dapat dijelaskan oleh waktu (minggu ke-n). Ini menunjukkan bahwa terdapat korelasi yang cukup kuat antara waktu dan harga, meskipun masih terdapat 37.88% variasi yang tidak dijelaskan oleh model (kemungkinan berasal dari fluktuasi pasar, berita, atau faktor fundamental).

Model Regresi Linier

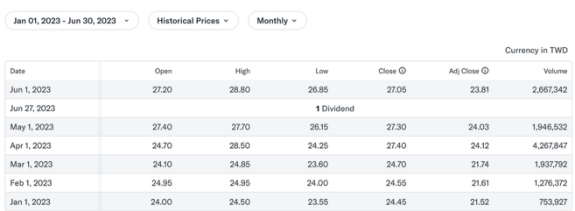
$$\hat{y} = 25.3586 - 0.1784\,x$$

Interpretasi;

Intercept (β_0) = 25.3566, perkiraan harga penutupan pada minggu ke-0 (akhir Maret 2023).

Slope (β_1) = 0.1784, rata-rata kenaikan harga sekitar 0.1784 TWD per minggu menunjukkan tren positif yang stabil selama periode tersebut.

4.3 Data Bulanan



Jan 01, 2023 - Jun 30, 2023 | Historical Prices | Monthly

Currency in TWD

Date	Open	High	Low	Close	Adj Close	Volume
Jun 1, 2023	27.20	28.80	26.85	27.05	23.81	2,667,342
Jun 27, 2023			1 Dividend			
May 1, 2023	27.40	27.70	26.15	27.30	24.03	1,946,532
Apr 1, 2023	24.70	28.50	24.25	27.40	24.12	4,267,847
Mar 1, 2023	24.10	24.85	23.60	24.70	21.74	1,937,792
Feb 1, 2023	24.95	24.95	24.00	24.55	21.61	1,276,372
Jan 1, 2023	24.00	24.50	23.55	24.45	21.52	753,927

Gambar 6. Data Periode Bulan Januari – Juni 2023 (Yahoo Finance, 2023)

MSE = 0.4343

Nilai ini menunjukkan bahwa rata-rata kuadrat kesalahan prediksi model hanya sekitar 0.43 TWD², yang tergolong sangat kecil mengingat harga saham berada dalam kisaran 24–27 TWD. Artinya, prediksi model linier sangat mendekati nilai aktual.

$R^2 = 0.9492$

94.92% variasi harga saham bulanan yang menunjukkan kecocokan yang sangat baik. Hanya sekitar 5% variasi yang tidak ditangkap oleh model.

Model Regresi Linier

$$\hat{y} = 28.2875 - 0.0932\,x$$

Interpretasi;

Intercept (β_0) = 24.1800, Estimasi harga saham pada bulan ke-0, yaitu Januari 2023.

Slope (β_1) = 0.5744, Rata-rata harga saham naik sekitar 0.57 TWD setiap bulan, menunjukkan tren positif selama periode analisis.

4.4 Data Akumulasi

Dalam persamaan tersebut, $x$adalah indeks bulan (dimulai dari 0 untuk Januari 2023), 0.5083adalah slope (kemiringan), dan 23.9220 adalah intercept. Artinya, harga penutupan saham diperkirakan meningkat sebesar 0.5083 TWD setiap bulan. Dengan demikian, model ini menangkap adanya tren kenaikan harga saham yang konsisten dan positif selama periode pengamatan. Dari evaluasi model, diperoleh MSE sebesar 0.3141, yang menunjukkan bahwa rata-rata kuadrat dari selisih antara prediksi model dan harga aktual hanya sekitar 0.31 TWD². Ini merupakan indikasi bahwa model memiliki performa prediksi yang cukup baik dalam konteks harga saham yang berada di kisaran 24–28 TWD. Selain itu, Koefisien Determinasi (R^2) yang mencapai 0.9455 menandakan bahwa sekitar 94.55% variasi harga saham dapat dijelaskan oleh model ini. Dengan kata lain, hubungan linier antara waktu dan harga penutupan sangat kuat dalam periode yang dianalisis.

Secara keseluruhan, model regresi linier sederhana ini memberikan hasil yang sangat baik untuk menangkap tren harga saham

P-ISSN :2580-4316
E-ISSN :2654-8054

bulanan 2012.TW selama enam bulan pertama tahun 2023. Tingginya nilai R^2 menunjukkan bahwa waktu merupakan prediktor yang sangat baik untuk menjelaskan variasi harga dalam jangka pendek ini. Meskipun model ini bersifat sederhana, ia memberikan baseline yang kuat dan dapat digunakan sebagai titik awal dalam pengembangan model prediktif yang lebih kompleks jika dibutuhkan, seperti regresi polinomial atau model berbasis neural network.

4.5 Analisis Actual Vs Predicted Close Prices

Gambar 7. Grafik Acctual vs Predictions Periode Bulan Januari – Juni 2023

Gambar 7., menunjukkan grafik perbandingan antara nilai aktual dan nilai prediksi harga penutupan saham (Close Prices) yang dihasilkan oleh model regresi (kemungkinan besar regresi linier atau model supervised lainnya). Pada sumbu horizontal (x-axis) ditampilkan indeks data (yang mewakili urutan waktu atau entri data), sedangkan sumbu vertikal (y-axis) menunjukkan nilai harga penutupan saham yang telah dinormalisasi atau diskalakan.

Secara visual, kurva harga aktual (berwarna biru) dan kurva harga prediksi (berwarna merah putus-putus) tampak sangat berdekatan dan mengikuti pola yang serupa. Hal ini menunjukkan bahwa model memiliki kapasitas prediksi yang tinggi, karena mampu menyesuaikan pola naik-turun harga saham dengan sangat baik. Hampir di setiap titik, garis prediksi mengikuti arah perubahan nilai aktual, termasuk pada puncak-puncak tajam (peak) dan lembah-lembah yang curam (trough). Ini mengindikasikan bahwa model tidak hanya mampu menangkap tren umum, tetapi juga cukup responsif terhadap fluktuasi jangka pendek dalam data.

Meskipun terdapat sedikit perbedaan pada beberapa titik ekstrem (misalnya pada lonjakan tertinggi di sekitar indeks ke-17), secara keseluruhan selisih antara nilai aktual dan prediksi sangat kecil, yang secara statistik dapat diterjemahkan ke dalam nilai MSE yang rendah dan Koefisien Determinasi (R^2) yang tinggi. Grafik ini menjadi bukti visual bahwa model yang digunakan memiliki performa prediktif yang sangat baik terhadap data yang diberikan.

Secara umum, grafik ini mencerminkan bahwa model yang dibangun memiliki akurasi tinggi dan generalisasi yang baik terhadap data aktual, menjadikannya layak digunakan untuk analisis tren harga atau prediksi jangka pendek. Namun demikian, untuk prediksi di luar data yang sudah dikenal (out-of-sample forecasting), tetap diperlukan validasi tambahan dan potensi penggunaan model yang lebih kompleks jika dibutuhkan.

V Kesimpulan

Hasil penelitian menunjukkan bahwa implementasi *data preprocessing* berperan krusial dalam meningkatkan akurasi dan stabilitas model regresi linier untuk prediksi harga saham. Melalui tahapan pembersihan data, seleksi fitur, dan normalisasi, model mampu menghasilkan performa yang lebih baik dengan nilai R^2 dan MSE yang lebih optimal. Pada level harian, regresi linier hanya menjelaskan sebagian kecil variasi harga, namun pada data mingguan dan bulanan performa meningkat signifikan hingga mampu menangkap tren dengan akurasi tinggi. Hal ini menegaskan bahwa regresi linier lebih sesuai untuk analisis tren jangka pendek hingga menengah dibanding fluktuasi harian yang kompleks. Meskipun sederhana, regresi linier dapat dijadikan model dasar (baseline) untuk prediksi harga saham, terutama jika dipadukan dengan teknik preprocessing yang tepat. Penelitian ini juga menekankan pentingnya pengembangan model lanjutan, seperti regresi non-linier atau algoritma pembelajaran mesin, guna mengakomodasi karakteristik pasar yang dinamis dan non-linier. Dengan demikian, kontribusi utama penelitian ini adalah menunjukkan bahwa preprocessing tidak hanya bersifat teknis, melainkan juga strategis dalam membangun model prediksi saham yang lebih reliabel.

Daftar Pustaka

Adhy, D. R., Anwar, N., Maesaroh, S., & Hermawan, R. (2025). Machine Learning dan Internet of Things (IoT):

Implementasi Machine Learning dalam Internet of Things. In *Star Digital Publishing*. Star Digital Publishing. https://www.stardigitalpublishing.com/2025/03/machine-learning-dan-internet-of-things.html

Agustina, N., Januhari, N. N. U., Jana, P., Suryawan, I. K. D., Tentua, M. N., Lumba, E., Setyoningrum, N. G., Hardyanto, R. H., Syah, F., Anwar, N., Purwantara, I. M. A., Fairuzabadi, M., & Ciptadi, P. W. (2025). Pengantar Data Science: Teori, Teknik, dan Aplikasinya di Era Digital. In *Yashmedia*. Yashmedia. https://yashmedia.id/pengantar-data-science-teori-teknik-dan-aplikasinya-di-era-digital/

Fairuzabadi, M., Adytia, P., Wahyuni, Prastyo, P. H., Resha, M., Anwar, N., Intan, I., Kurniawan, M. R., Amiruddin, Wulan, N., Sekti, B. A., & Erna, A. (2024). Machine Learning: Konsep, Algoritma dan Implementasi. In *Kita Menulis*. Yayasan Kita Menulis. https://kitamenulis.id/2024/11/29/machine-learning-konsep-algoritma-dan-implementasi/

Gupta, S., & Dhingra, B. (2019). Feature selection and dimensionality reduction techniques for stock market prediction: A survey. *International Journal of Computer Applications*, *178*(7), 1–6. https://doi.org/10.5120/ijca2019918756

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques. In 3 (Ed.), *Morgan Kaufmann.* Morgan Kaufmann.

Jain, R., Kumar, A., & Singhal, A. (2018). Impact of normalization on the performance of data mining algorithms. *International Journal of Computer Applications*, *181*(7), 28–34. https://doi.org/10.5120/ijca2018917656

Mirfan, SAS, A., Indrayani, L., Erzed, N., Anwar, N., Ningsih, S. R., Stephane, I., Sekti, B. A., Simarmata, J., & Lubis, M. (2024). Riset Teknologi Informasi. In *Kita Menulis*. Yayasan Kita Menulis. https://kitamenulis.id/2024/10/07/riset-teknologi-informasi/

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259–268. https://doi.org/10.1016/j.eswa.2014.07.040

Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, *50*(1), 258–269. https://doi.org/10.1016/j.dss.2010.08.028

Yahoo Finance. (2023). Taishin Financial Holdings Co., Ltd. (2012.TW) Monthly Historical Data}, Available:https://finance.yahoo.com/quote/2012.TW/history?frequency=1mo.