

Evaluasi Komparatif Model Regresi Prediksi Saham BBKA dengan Analitik dan Visualisasi Interaktif Menggunakan Streamlit

Muhammad Hilmy Munsarif, Amiq Fahmi*, Agus Winarno

Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Kota Semarang, Indonesia
Email: ¹112202206821@mhs.dinus.ac.id, ^{2,*}amiq.fahmi@dsn.dinus.ac.id, ³agus.winarno@dsn.dinus.ac.id

Email Penulis Korespondensi: amiq.fahmi@dsn.dinus.ac.id

Submitted: 06/10/2025; Accepted: 11/12/2025; Published: 31/12/2025

Abstrak—Ketidakstabilan harga saham yang dipengaruhi oleh faktor fundamental dan teknikal menimbulkan kompleksitas dalam proses prediksi serta menuntut model yang akurat dan mampu melakukan generalisasi. Penelitian sebelumnya masih berfokus pada regresi linier konvensional tanpa membandingkannya dengan metode regularisasi maupun menelaah kontribusi rekayasa fitur dalam meningkatkan performa prediktif. Kontribusi penelitian ini adalah mengevaluasi dan membandingkan efektivitas Regresi Linier, Ridge Regression, dan Lasso Regression dalam memprediksi harga penutupan saham PT Bank Central Asia Tbk (BBKA). Kebaruan penelitian ini terletak pada penerapan tahapan preprocessing dan feature engineering yang menghasilkan tujuh variabel turunan, yaitu daily range, open–close change, daily return, lag features, moving average, volatility, dan transformasi logaritmik. Evaluasi model menggunakan Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2). Hasil utama menunjukkan bahwa Regresi Linier memiliki akurasi tinggi pada data pelatihan namun mengalami overfitting pada data uji, Ridge Regression tidak memberikan peningkatan stabilitas yang berarti, sedangkan Lasso Regression menjadi model paling stabil dengan nilai R^2 sebesar 0,8247. Temuan ini memberikan manfaat berupa dasar pemilihan metode prediksi yang lebih stabil dan akurat untuk digunakan dalam analisis harga saham dengan volatilitas tinggi.

Kata Kunci: Prediksi Harga Saham; Regresi Linier; Ridge Regression; Lasso Regression; Feature Engineering; Evaluasi Model

Abstract—The instability of stock prices influenced by fundamental and technical factors creates complexity in the prediction process and requires models that are both accurate and capable of generalizing well. The contribution of this study is to evaluate and compare the performance of Linear Regression, Ridge Regression, and Lasso Regression in predicting the closing price of PT Bank Central Asia Tbk (BBKA). The novelty of this research lies in the application of preprocessing and feature engineering procedures that generate seven derived variables, namely daily range, open–close change, daily return, lag features, moving average, volatility, and logarithmic transformation. Model evaluation was conducted using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The main findings indicate that Linear Regression shows high accuracy on the training data but suffers from overfitting on the test data, Ridge Regression does not provide a significant improvement in stability, while Lasso Regression demonstrates the most stable performance with an R^2 of 0.8247. These findings offer practical benefits by providing a more stable and accurate predictive method for analyzing stock prices with high volatility.

Keywords: Stock Price Prediction; Linear Regression; Ridge Regression; Lasso Regression; Feature Engineering; Model Evaluation

1. PENDAHULUAN

Pasar modal, baik secara global maupun nasional, memainkan peran sentral dalam mendorong pertumbuhan ekonomi melalui mekanisme pembiayaan dan investasi yang saling melengkapi [1]. Di Indonesia, saham merupakan instrumen utama yang menawarkan potensi keuntungan melalui dividen dan *capital gain*. Namun, pergerakan harga saham sangat dipengaruhi oleh faktor fundamental, teknikal, dan makroekonomi seperti PDB, inflasi, suku bunga, serta stabilitas politik. Tingginya volatilitas ini menimbulkan tantangan dalam pengambilan keputusan investasi berbasis data, sehingga diperlukan pendekatan prediktif yang lebih reliabel dan tidak hanya bergantung pada pola jangka pendek [2].

Dalam lanskap pasca-pandemi dan digitalisasi keuangan, kebutuhan akan model prediksi saham yang akurat semakin meningkat. PT Bank Central Asia Tbk (BBKA) menonjol sebagai perusahaan dengan kapitalisasi pasar tinggi dan tingkat likuiditas kuat, menjadikannya objek analisis yang relevan. Berbagai pendekatan telah diterapkan, seperti regresi *linier*, *polynomial regression*, dan *artificial neural network* (ANN) [3], [4]. Namun, metode-metode tersebut belum sepenuhnya mampu mengatasi karakteristik data yang non-linier dan multikolinear sehingga menghasilkan prediksi yang kurang stabil [5].

Keterbatasan regresi linier dalam menghadapi overfitting dan rendahnya kemampuan generalisasi mendorong penerapan metode regularisasi. Pendekatan *Ridge* dan *Lasso Regression* menawarkan kemampuan penekanan kompleksitas model dan seleksi fitur yang lebih efisien, sehingga meningkatkan stabilitas prediksi dalam konteks data keuangan [6]. Temuan empiris juga menunjukkan bahwa regularisasi mampu meningkatkan akurasi dan menurunkan variansi model pada dataset berskala besar dan dinamis [7].

Seiring meningkatnya kebutuhan interpretasi hasil analisis, integrasi model prediktif ke dalam platform interaktif seperti Streamlit menjadi semakin relevan. Streamlit memungkinkan penyajian hasil prediksi secara

visual dan eksploratif, sehingga meningkatkan kemudahan interpretasi model bagi pengguna non-teknis [8]. Di Indonesia, penerapan dashboard interaktif untuk analisis saham masih terbatas meskipun berpotensi memperkuat proses pengambilan keputusan berbasis visualisasi [9].

Penelitian terdahulu belum melakukan komparasi langsung antara *Linear Regression*, *Ridge Regression*, dan *Lasso Regression* pada saham BBCA periode 2023–2024, serta belum mengkaji kontribusi *feature engineering* dalam meningkatkan akurasi prediksi. Selain itu, belum ada penelitian yang mengintegrasikan model regresi tersebut ke dalam platform visual interaktif berbasis *Streamlit* [10], [11]. Berdasarkan celah penelitian tersebut, novelty studi ini terletak pada kombinasi evaluasi komparatif tiga metode regresi, pemanfaatan variabel turunan untuk memperkaya representasi data, dan implementasi model terbaik dalam aplikasi interaktif. Rumusan masalah penelitian ini meliputi: (1) bagaimana perbandingan performa *Linear*, *Ridge*, dan *Lasso Regression* dalam memprediksi harga penutupan saham BBCA periode 2023–2024, (2) model mana yang menghasilkan akurasi dan stabilitas terbaik berdasarkan *RMSE*, *MAE*, dan R^2 dan (3) bagaimana integrasi model melalui *Streamlit* dapat mendukung interpretasi hasil prediksi secara visual dan aplikatif.

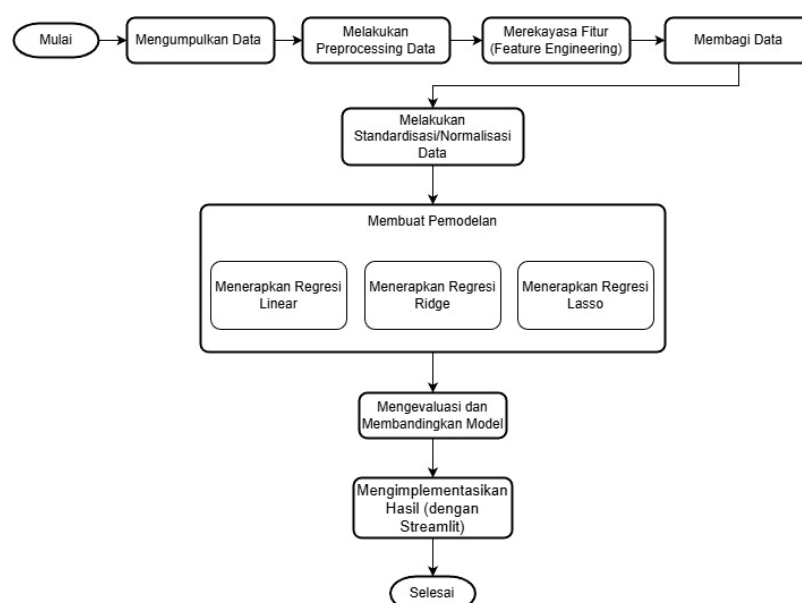
2. METODOLOGI PENELITIAN

Tahapan penelitian ini terdiri atas rangkaian proses terstruktur yang dirancang untuk menghasilkan model prediksi harga saham BBCA yang akurat serta dapat diimplementasikan ke dalam aplikasi interaktif. Proses dimulai dengan pengumpulan data historis saham BBCA periode 2023–2024 dari Yahoo Finance dalam format CSV sebagai dasar analisis [10]. Tahap berikutnya adalah melakukan preprocessing data, yang mencakup pemeriksaan nilai hilang, penyesuaian format tanggal, penerapan transformasi logaritmik pada variabel dengan distribusi tidak simetris, serta standarisasi menggunakan *StandardScaler* untuk menyetarakan skala variabel numerik [12].

Setelah itu dilakukan rekayasa fitur (*feature engineering*) dengan membangun variabel turunan berbasis volatilitas, tren, dan pola temporal, seperti *lag*, *moving average*, dan *return* harian guna meningkatkan kemampuan model dalam menangkap dinamika pergerakan harga saham [10], [13].

Tahap selanjutnya adalah membagi data menggunakan pendekatan *train-test split* secara kronologis sesuai karakteristik deret waktu guna mencegah *data leakage* dan menghasilkan evaluasi yang lebih representative [14]. Sebelum model dibangun, dilakukan standarisasi atau normalisasi data menggunakan *StandardScaler*, yang diterapkan pada data latih dan kemudian diproyeksikan ke data uji. Setelah pembagian data, dilakukan proses pemodelan yang mencakup penerapan tiga algoritma regresi—*Linear Regression*, *Ridge Regression*, dan *Lasso Regression*—dengan konfigurasi parameter yang konsisten agar perbandingan performa bersifat objektif [3].

Tahap berikutnya adalah mengevaluasi dan membandingkan model menggunakan metrik *MAE*, *RMSE*, dan R^2 untuk menentukan model dengan performa terbaik [4]. Tahap akhir adalah mengimplementasikan hasil ke dalam aplikasi *Streamlit* sehingga prediksi dapat disajikan secara interaktif melalui visualisasi yang informatif [8], [15], [16]. Alur lengkap proses penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Mengumpulkan data

Tahap awal penelitian dilakukan dengan mengumpulkan data historis saham PT Bank Central Asia Tbk (BBCA) melalui platform *Yahoo Finance* dalam format *comma-separated values* (CSV). Dataset mencakup periode perdagangan harian tahun 2023–2024 dan berisi enam atribut utama, yaitu tanggal transaksi (*date*), harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), harga penutupan (*close*), dan volume perdagangan (*volume*). Harga penutupan dipilih sebagai variabel target karena dianggap paling merepresentasikan nilai akhir saham dalam satu sesi perdagangan, sedangkan atribut lainnya berfungsi sebagai variabel awal untuk proses rekayasa fitur dan pemodelan [10], [13].

Proses pengambilan data dilakukan menggunakan bahasa pemrograman Python melalui pustaka *yfinance* untuk mengunduh data langsung dari *Yahoo Finance*, dan *pandas* untuk membaca serta memproses dataset ke dalam bentuk *dataframe*. Pemeriksaan awal dilakukan terhadap jumlah hari perdagangan, kelengkapan atribut, konsistensi format tanggal, serta potensi *missing values* yang umumnya muncul akibat hari libur bursa atau ketidaksesuaian pencatatan. Jika ditemukan data yang hilang, penanganan dilakukan dengan metode imputasi sederhana atau penghapusan baris guna memastikan kualitas dataset tetap terjaga sebelum memasuki tahap praproses (*preprocessing*).

Rentang data dua tahun dipilih untuk menangkap variasi harga saham dalam skala jangka pendek hingga menengah secara representatif. Pemilihan BBCA sebagai objek penelitian didasarkan pada kapitalisasi pasar yang besar, tingkat likuiditas tinggi, serta pengaruh signifikan terhadap pergerakan Indeks Harga Saham Gabungan (IHSG) [11]. Variabel *date* berfungsi sebagai penanda urutan deret waktu dan tidak dilibatkan sebagai prediktor dalam pemodelan.

Tabel 1. Jenis jenis variabel

Variabel	Jenis	Penjelasan
Date	Atribut waktu	Tanggal dilaksanakannya transaksi saham.
Open	Independen	Nilai harga saham pada saat awal perdagangan dimulai.
High	Independen	Harga naik saham yang dicapai selama satu hari aktivitas perdagangan.
Low	Independen	Harga turun saham yang tercatat pada hari aktivitas perdagangan tersebut.
Close	Dependen (target)	Nilai penutupan saham pada akhir sesi perdagangan harian.
Volume	Independen	Jumlah total saham yang diperjualbelikan dalam satu hari perdagangan.

Tahap pengumpulan data ini diikuti dengan proses eksplorasi awal menggunakan pustaka *pandas*, *numpy*, dan *matplotlib* untuk memeriksa struktur data, mendeteksi pola dasar, serta menilai kebutuhan terhadap praproses dan rekayasa fitur. Hasil dari tahap ini berupa dataset yang telah tervalidasi dan siap diolah lebih lanjut dalam proses *preprocessing* dan *feature engineering*, yang akan dibahas pada bagian selanjutnya.

2.2 Melakukan Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan bahwa data historis saham PT Bank Central Asia Tbk (BBCA) dalam kondisi layak digunakan sebelum memasuki proses rekayasa fitur dan pemodelan [12]. Prosedur pada tahap ini mencakup pemeriksaan nilai hilang, validasi format tanggal, penyusunan ulang urutan data, serta pemeriksaan distribusi awal sebagai bagian dari eksplorasi kualitas data.

1. Penanganan Nilai Hilang (*Missing Values*)

Pemeriksaan nilai hilang dilakukan terhadap seluruh variabel (*open*, *high*, *low*, *close*, *volume*, dan *date*) untuk memastikan tidak ada entri yang hilang akibat hari libur bursa atau kesalahan pencatatan. Hasil pemeriksaan menunjukkan bahwa dataset tidak mengandung *missing value* sehingga tidak diperlukan tindakan imputasi maupun penghapusan baris. Langkah ini dilakukan karena tahapan pengecekan kelengkapan data merupakan prosedur standar dalam pengolahan data deret waktu [12].

2. Validasi Format dan Struktur Data

Setelah validasi, kolom tersebut dijadikan indeks agar urutan data *time-series* terjaga. Penyusunan ulang urutan data dilakukan berdasarkan tanggal secara menaik untuk memastikan tidak terjadi ketidakkonsistenan urutan. Seluruh kolom numerik juga diperiksa tipe datanya untuk menjamin bahwa data dapat diproses secara matematis tanpa kesalahan [14].

3. Pemeriksaan *Outlier* dan Distribusi Awal

Eksplorasi distribusi awal dilakukan menggunakan statistik deskriptif dan visualisasi sederhana guna mengidentifikasi keberadaan nilai ekstrem atau ketidakwajaran pola data. Hasil eksplorasi menunjukkan adanya skewness pada volume perdagangan, tetapi penanganan transformasi distribusi tidak dilakukan pada tahap preprocessing, melainkan pada tahap *feature engineering* karena berkaitan langsung dengan pembuatan variabel turunan [17].

2.3 Merekayasa Fitur (Feature Engineering)

Tahap feature engineering dilakukan untuk membentuk variabel turunan dari data mentah dengan tujuan meningkatkan kemampuan model dalam menangkap pola pergerakan harga saham yang bersifat non-linier, autokorelatif, dan volatil [10], [13]. Data awal yang diperoleh dari *Yahoo Finance*, yang mencakup harga Open, High, Low, Close, dan Volume diolah menjadi sebelas fitur prediktor yang digunakan dalam proses pemodelan.

Fitur-fitur tersebut mencerminkan tiga dimensi penting dalam analisis pasar saham, yaitu volatilitas harian, arah tren, dan dinamika temporal. Indikator volatilitas direpresentasikan melalui *daily range* dan *volatility_5*, sedangkan dinamika tren diperoleh dari moving average (MA_5 dan MA_20). Pola autokorelasi jangka pendek ditangkap melalui fitur lag seperti *Close_Lag1*, *Close_Lag2*, dan *Volume_Lag1*. Selain itu, transformasi logaritmik diterapkan pada variabel Volume untuk mengurangi skewness distribusi data. Adapun variabel target dalam penelitian ini adalah *log_Close*, yang merupakan transformasi logaritmik dari harga penutupan. Tabel 2 berikut menyajikan daftar fitur prediktor beserta rumus perhitungan dan penjelasannya.

Tabel 2. Jenis Feature Engineering

Nama Fitur	Rumus / Formula	Penjelasan
Daily Range	$Daily\ Range_t = High_t - Low_t$	Selisih harga tertinggi dan terendah tiap hari
Open-Close Change	$Open - Close\ Change_t = Close_t - Open_t$	Perbedaan antara harga penutupan dan pembukaan
Daily Return (%)	$Daily\ Return_t = \frac{Close_t - Close_{t-1}}{Close_{t-1}} \times 100\%$	Persentase perubahan harga penutupan dibandingkan hari sebelumnya
Close Lag1	$Close_Lag1_t = Close_{t-1}$	Harga penutupan hari sebelumnya
Close Lag2	$Close_Lag2_t = Close_{t-2}$	Harga penutupan dua hari sebelumnya
Volume Lag1	$Volume_Lag1 = Volume_t - 1$	Volume perdagangan hari sebelumnya
MA5	$MA_5(t) = \frac{1}{k} \sum_{i=0}^{k-1} Close_{t-i}$	Rata-rata bergerak 5 hari untuk menangkap tren jangka pendek
MA20	$MA_{20}(t) = \frac{1}{k} \sum_{i=0}^{k-1} Close_{t-i}$	Rata-rata bergerak 20 hari untuk menangkap tren jangka menengah
Volatility_5	$Volatility_t = \sqrt{\frac{1}{n-1} \sum_{i=0}^n (Return_{t-i} - \overline{Return})^2}$	Standar deviasi return 5 hari terakhir, sebagai indikator volatilitas
log_Close	$Log_Close_t = \ln(Close_t)$	Transformasi logaritmik harga penutupan
log_Volume	$Log_Volume_t = \ln(Volume_t)$	Transformasi logaritmik volume perdagangan

Penelitian-penelitian terdahulu menegaskan bahwa pembentukan fitur berbasis teknikal dan temporal dapat meningkatkan akurasi model prediksi harga saham. Studi dalam [10] menyoroti kontribusi fitur berbasis autokorelasi terhadap performa model regresi, sementara [13], [18] menunjukkan bahwa indikator teknikal seperti moving average dan volatilitas memiliki peran signifikan dalam model prediksi harga saham berbasis machine learning. Dengan demikian, keseluruhan proses feature engineering ini dirancang untuk memastikan bahwa model memperoleh informasi yang representatif terhadap dinamika harga saham BBKA.

2.4 Membagi Data

Setelah tahap *feature engineering* selesai, data historis saham BBKA dipisahkan menjadi dua bagian: data latih (*training set*) dan data uji (*testing set*). Tujuan pemisahan ini adalah agar model mempelajari pola dari data historis, lalu dievaluasi menggunakan data baru yang belum pernah digunakan dalam pelatihan [10].

Sehubungan dengan data yang tersusun dalam urutan waktu (*time series*), pembagian dilakukan secara kronologis untuk mencegah data leakage, yaitu masuknya informasi masa depan ke dalam proses pelatihan [18]. Pembagian dataset dilakukan dengan 80% untuk pelatihan dan 20% untuk pengujian dipandang ideal, karena cukup untuk membangun model yang stabil sekaligus menilai performa secara objektif [19].

Pada proses ini, data tidak melalui prosedur *shuffling* karena akan merusak urutan temporal yang menjadi karakteristik utama data *time series* [13]. Penelitian ini juga tidak menggunakan metode *TimeSeriesSplit*, karena pendekatan train-test split kronologis dianggap lebih sesuai dengan tujuan prediksi satu langkah ke depan (*one-*

step ahead forecasting) sebagaimana diterapkan pada data saham harian [11]. Pendekatan ini memastikan bahwa seluruh fitur yang telah direkayasa digunakan secara konsisten sebagai prediktor, sedangkan variabel target tetap berupa harga penutupan yang telah ditransformasikan [7].

Sejumlah penelitian sebelumnya menunjukkan bahwa strategi pembagian data secara berurutan menghasilkan performa evaluasi yang lebih realistis dibandingkan metode acak, khususnya pada analisis saham berbasis time series [10], [13], [20]. Dengan demikian, pendekatan ini dipilih untuk menjaga validitas serta reliabilitas model prediktif yang dikembangkan dalam penelitian ini [19].

2.5 Melakukan Standardisasi/Normalisasi Data

Tahap standardisasi dilakukan setelah proses pembagian data untuk memastikan bahwa seluruh transformasi statistik hanya dipelajari dari data latih dan tidak melibatkan informasi dari data uji. Pendekatan ini penting untuk mencegah data leakage, yaitu kondisi ketika parameter transformasi terbentuk dari keseluruhan data sehingga memengaruhi objektivitas evaluasi model [19].

Standardisasi diterapkan menggunakan teknik *StandardScaler*, yang mentransformasikan setiap fitur numerik ke dalam distribusi dengan rata-rata nol dan simpangan baku satu. Transformasi ini dinyatakan dengan persamaan (1).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

di mana μ adalah nilai rata-rata dan σ adalah simpangan baku dari masing-masing fitur. Teknik ini digunakan karena model Ridge dan Lasso bersifat sensitif terhadap skala fitur; perbedaan skala antarvariabel dapat menyebabkan penalti koefisien menjadi tidak proporsional sehingga memengaruhi stabilitas solusi [5], [6], [7].

Proses standardisasi tidak dilakukan secara manual pada dataset, tetapi diintegrasikan ke dalam *pipeline* bersama model regresi. Dengan demikian, *StandardScaler* akan melakukan *fit* hanya pada data latih dan menerapkan transformasi yang sama pada data uji. Pendekatan ini memastikan bahwa prosedur transformasi berlangsung konsisten dan mengikuti praktik terbaik dalam pemodelan berbasis machine learning [19], [20].

Hanya fitur prediktor (X) yang distandardisasi, sementara variabel target tetap dalam bentuk aslinya kecuali ketika transformasi logaritmik diterapkan pada tahap preprocessing. Sejumlah studi melaporkan bahwa standardisasi berbasis z-score meningkatkan stabilitas numerik dan mempercepat proses optimasi pada model regresi dengan regularisasi [13], [18].

2.6 Membuat Pemodelan

Pendekatan pemodelan prediktif dalam penelitian ini menggunakan regresi linier serta dua metode regularisasi, yaitu Ridge Regression dan Lasso Regression. Ketiga model dipilih karena dinilai efektif dalam meningkatkan stabilitas prediksi dan mengatasi risiko overfitting pada data historis saham yang fluktuatif [5], [6], [14]. Seluruh model dilatih menggunakan dataset yang sama dengan pembagian 80% data latih dan 20% data uji, di mana proses normalisasi hanya diterapkan pada data latih dan kemudian diproyeksikan ke data uji untuk mencegah terjadinya *data leakage* [7].

Penelitian ini secara sengaja tidak menggunakan cross-validation (CV) maupun *hyperparameter tuning*. Pada data *time series*, penggunaan CV dinilai kurang tepat karena prosedur pembagian data secara acak dapat menyebabkan informasi masa depan masuk ke dalam data latih (*look-ahead bias*) [14], [20]. Oleh karena itu, pemodelan dilakukan menggunakan pembagian kronologis 80:20 yang lebih sesuai untuk struktur deret waktu.

Selain itu, nilai hyperparameter α pada Ridge dan Lasso tidak dituning, melainkan ditetapkan secara manual (Ridge = 1.0 dan Lasso = 0.001) agar hasil prediksi stabil dan konsisten dengan implementasi aktual. Pendekatan ini sejalan dengan penelitian finansial sebelumnya yang menerapkan fixed-alpha untuk menjaga reproduktibilitas model serta menghindari bias akibat optimasi berlebihan [3], [10], [11].

Dengan demikian, pendekatan ini menjaga integritas metodologis serta memungkinkan hasil pemodelan direplikasi secara transparan.

1) Regresi Linier

Regresi linier merupakan salah satu metode statistik konvensional yang sederhana, tetapi tetap banyak digunakan hingga sekarang untuk menjelaskan hubungan antara variabel dependen dengan satu maupun lebih variabel independen [10]. fungsi objektif Regresi Linier dapat dituliskan seperti (2).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (2)$$

Pada persamaan tersebut, y merepresentasikan variabel dependen (misalnya *close*), sedangkan, $x_1, x_2, \dots, x_n$ adalah variabel independen seperti *Open*, *High*, *Low*, *Volume*, maupun variabel turunan lainnya. Sementara itu, β_0 merupakan *intercept*, β_i menyatakan koefisien regresi, ε adalah komponen *error*.

Meskipun regresi linier tergolong mudah diterapkan dan diinterpretasikan, metode ini rentan terhadap *overfitting*, terutama ketika jumlah prediktor cukup banyak atau terdapat multikolinearitas antar variabel. Untuk mengatasi kendala tersebut, pendekatan regularisasi seperti *Ridge* dan *Lasso Regression* digunakan sebagai solusi yang efektif dalam menstabilkan model dan meningkatkan generalisasi [5], [14].

2) Ridge Regression (Regularisasi L2)

Ridge Regression merupakan pengembangan dari regresi linier dengan menambahkan penalti berupa kuadrat koefisien regresi. Tujuannya adalah mengecilkan nilai koefisien untuk mengurangi varians model dan mengatasi *multikolinearitas* [5], [6]. Fungsi objektif *Ridge* dituliskan menggunakan formula (3).

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (3)$$

Parameter λ mengontrol besar kecilnya penalti. Pada penelitian ini, *Ridge Regression* diterapkan menggunakan nilai $\alpha = 1.0$, yang memberikan tingkat regularisasi moderat dan mampu menjaga keseimbangan antara bias dan varians. Nilai tersebut efektif untuk data keuangan yang memiliki korelasi tinggi antar fitur [20], [21].

3) Lasso Regression (Regularisasi L1)

Lasso (Least Absolute Shrinkage and Selection Operator) menggunakan penalti berupa nilai absolut koefisien regresi. Selain mengurangi varians model, *Lasso* juga dapat melakukan seleksi fitur dengan mengecilkan sebagian koefisien menjadi nol. Fungsi objektif *Lasso* menggunakan formula (4).

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (4)$$

penelitian ini, *Lasso Regression* menggunakan $\alpha = 0.001$ dengan jumlah iterasi maksimum 10.000 untuk memastikan proses konvergensi berjalan optimal. Penggunaan nilai penalti yang rendah dipilih agar *Lasso* tidak menghilangkan terlalu banyak fitur penting, mengingat data saham memiliki karakteristik volatilitas tinggi dan risiko kehilangan informasi jika regularisasi terlalu kuat. Temuan penelitian finansial sebelumnya juga menunjukkan bahwa *Lasso* dengan penalti rendah mampu meningkatkan kinerja prediktif pada data dengan *multikolinearitas* tinggi [7], [21].

2.7 Mengevaluasi dan Membandingkan Model

Proses evaluasi model memiliki peranan penting dalam menentukan seberapa baik algoritma regresi bekerja pada penelitian ini. Beberapa metrik yang digunakan sebagai tolok ukur mencakup *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan Koefisien Determinasi (R^2).

a. Root Mean Squared Error (RMSE).

Root Mean Squared Error (RMSE) merupakan salah satu metrik evaluasi yang umum digunakan untuk mengukur tingkat kesalahan prediksi dalam model regresi. RMSE dihitung dengan mengambil akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan nilai prediksi, sebagaimana ditunjukkan pada formula (5).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

Semakin kecil nilai RMSE, maka semakin rendah pula tingkat kesalahan prediksi model secara keseluruhan. Karena bersifat sensitif terhadap adanya error yang besar (*outlier*), metrik ini sering dipakai sebagai indikator utama dalam menilai performa model regresi [19].

b. Mean Absolute Error (MAE).

Mean Absolute Error (MAE) merupakan salah satu metrik evaluasi yang digunakan untuk mengukur akurasi model regresi dengan menghitung rata-rata selisih absolut antara nilai aktual dan nilai prediksi. Formula MAE ditunjukkan pada rumus (6).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Kelebihan utama dari *Mean Absolute Error* (MAE) terletak pada kemudahan interpretasinya, karena nilai kesalahan yang dihasilkan berada dalam skala yang sama dengan data asli. Hal ini menjadikan MAE lebih intuitif dibandingkan RMSE dalam menilai seberapa besar deviasi antara hasil prediksi dan nilai aktual [22]. MAE juga tidak terlalu sensitif terhadap *outlier*, sehingga cocok digunakan dalam konteks evaluasi model yang mengutamakan stabilitas prediksi secara umum.

c. Koefisien Determinasi (R^2).

Koefisien determinasi R^2 digunakan untuk menilai sejauh mana variasi pada variabel dependen dapat dijelaskan oleh model regresi (7).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Nilai R^2 yang mendekati 1 mengindikasikan bahwa model mampu menjelaskan sebagian besar variasi data dengan baik. Sebaliknya, jika nilainya semakin dekat ke 0, maka model dianggap kurang representatif dalam menggambarkan pola data [23].

Dalam penelitian ini, model Regresi Linier, Ridge, dan Lasso dievaluasi menggunakan tiga indikator utama, yaitu *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi (R^2). Penetapan

model terbaik didasarkan pada kombinasi nilai RMSE dan MAE yang lebih rendah serta nilai R^2 yang lebih tinggi, sehingga tidak hanya menggambarkan akurasi prediksi, tetapi juga kemampuan model dalam menjelaskan variabilitas data secara lebih komprehensif. Pendekatan evaluasi multimetrik ini selaras dengan temuan penelitian terdahulu yang menegaskan bahwa penggunaan satu metrik saja berpotensi menghasilkan penilaian yang bias, sehingga diperlukan lebih dari satu indikator untuk memperoleh gambaran kinerja model yang lebih menyeluruh [19], [23].

Sebagai bagian dari proses regularisasi, Lasso Regression berfungsi sekaligus sebagai mekanisme seleksi fitur melalui pengecilan koefisien tertentu menjadi nol. Hasil analisis menunjukkan bahwa Lasso mempertahankan empat fitur yang paling relevan dalam memprediksi harga saham BBKA, yaitu Daily Return (%), Close Lag1, Volume Lag1, dan MA5. Keempat fitur tersebut memiliki koefisien non-nol sehingga dianggap memberikan kontribusi signifikan terhadap model. Sebaliknya, fitur seperti Volume, Daily Range, Open–Close Change, Close Lag2, MA20, Volatility_5, serta log_Volume mengalami pengecilan koefisien menjadi nol, sehingga dieliminasi karena kontribusinya tidak cukup kuat. Temuan ini konsisten dengan literatur mengenai efektivitas regularisasi L1 dalam menyaring variabel yang paling informatif, mengurangi kompleksitas model, serta mengurangi risiko overfitting pada data finansial yang umumnya memiliki korelasi tinggi antar fitur [5], [6], [7], [14].

Dengan demikian, hasil seleksi fitur oleh Lasso memberikan landasan empiris yang lebih kuat dalam membangun model prediksi harga saham yang stabil, efisien, dan mampu menangkap pola signifikan pada data yang bersifat dinamis dan kompleks.

2.8 Mengimplementasikan Hasil (dengan Streamlit)

Tahap akhir penelitian ini adalah mengimplementasikan model prediktif ke dalam aplikasi interaktif berbasis Streamlit. Implementasi ini bertujuan agar hasil pemodelan tidak hanya disajikan dalam bentuk analisis statistik, tetapi juga dapat diakses, divisualisasikan, serta diuji secara langsung oleh pengguna. Streamlit dipilih karena mampu menghasilkan antarmuka analitis yang ringan, responsif, dan mudah dioperasikan, termasuk oleh pengguna non-teknis [16].

Aplikasi yang dibangun memuat serangkaian komponen utama, yaitu:

- Modul pemuatan model, di mana model regresi terbaik dimuat dari file .joblib menggunakan pendekatan wrapper. Pendekatan ini memastikan seluruh proses preprocessing, seperti standarisasi fitur dan transformasi logaritmik, tetap konsisten antara fase pelatihan dan fase prediksi.
- Form input fitur, yang memungkinkan pengguna memasukkan nilai-nilai feature engineering seperti *Daily Range*, *Open–Close Change*, *Daily Return*, *fitur lag*, *moving average*, volatilitas berbasis *rolling window*, serta fitur *logaritmik*. Seluruh input ini digunakan untuk menghasilkan prediksi harga penutupan.
- Tampilan hasil prediksi, yang menyajikan estimasi harga penutupan saham BBKA berdasarkan nilai fitur yang dimasukkan oleh pengguna. Metrik evaluasi seperti *RMSE*, *MAE*, dan R^2 hanya dijelaskan pada tahap analisis performa dalam artikel, bukan ditampilkan di aplikasi.
- Visualisasi interaktif, berupa grafik perbandingan antara harga aktual dan harga prediksi. Grafik ini membantu pengguna memahami pola hubungan antara prediksi dan data historis secara intuitif melalui tampilan visual yang mudah dipahami.

Pendekatan berbasis dashboard ini telah diakui efektif dalam meningkatkan transparansi model machine learning dan mendukung interpretasi hasil oleh pengguna *non-teknis* [24]. Dengan demikian, integrasi Streamlit dalam penelitian ini memastikan bahwa model prediksi tidak hanya akurat secara kuantitatif, tetapi juga mudah dipahami dan dapat digunakan untuk mendukung proses pengambilan keputusan investasi secara praktis.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Analisis Data

Penelitian diawali dengan Exploratory Data Analysis (EDA) untuk memahami pola historis, struktur data, dan karakteristik utama saham PT Bank Central Asia Tbk. (BBKA) periode 2023–2024. EDA berfungsi sebagai tahap penting dalam memastikan integritas data dan mengarahkan proses feature engineering, sebagaimana direkomendasikan pada studi-studi sebelumnya mengenai analisis data pasar modal [1], [17].

Analisis deret waktu menunjukkan bahwa harga penutupan BBKA mengalami tren kenaikan yang stabil sejak awal 2023 hingga kuartal ketiga 2024, sebelum terjadi koreksi moderat di akhir periode. Sebaliknya, volume transaksi tampak jauh lebih volatil, dengan beberapa lonjakan signifikan yang berkaitan dengan sentimen pasar dan peristiwa eksternal. Pola ini konsisten dengan karakteristik saham blue-chip, di mana harga cenderung stabil namun volume lebih reaktif terhadap informasi jangka pendek [3].

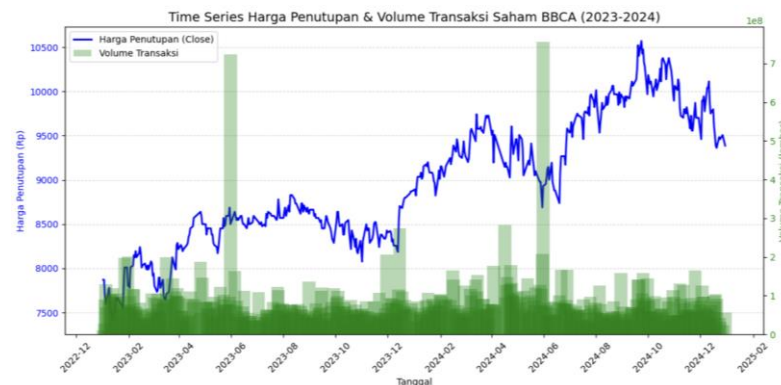
Temuan ini memperkuat dasar metodologis bahwa variabel harga dan volume sama-sama dibutuhkan dalam model prediksi, karena keduanya mewakili dinamika pasar dari perspektif tren dan reaktivitas [2].

Gambar 2 menunjukkan pergerakan time series harga penutupan dan volume. Tren harga yang cenderung smooth mengindikasikan adanya pola jangka menengah, sehingga teknik regresi berbasis regularisasi relevan digunakan untuk menghindari overfitting ketika menangani variabel harga yang saling berkorelasi sangat kuat [6],

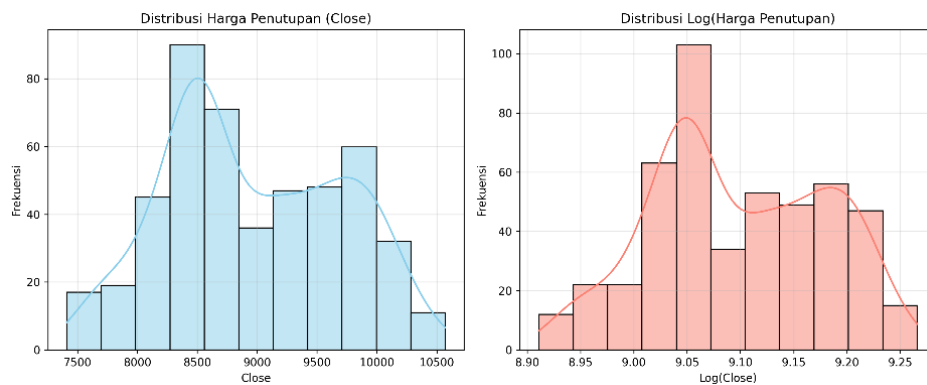
[21]. Di sisi lain, fluktuasi volume yang tajam menjadi sinyal bahwa feature engineering berbasis perubahan harian, moving average, dan volatilitas perlu dimasukkan untuk menangkap dinamika jangka pendek pasar [20].

Tahap selanjutnya adalah analisis distribusi harga penutupan (Close) dan hasil transformasi logaritmiknya (Log_Close), seperti terlihat pada Gambar 3. Distribusi asli condong ke kanan (right-skewed), yang kurang ideal bagi regresi linier karena varians tidak stabil dan error tidak berdistribusi mendekati normal. Transformasi logaritmik berhasil mengurangi skewness dan menstabilkan varians, sehingga lebih sesuai sebagai variabel target dalam pemodelan regresi. Transformasi log juga merupakan praktik umum pada prediksi harga saham untuk meningkatkan akurasi dan stabilitas model [11], [12].

Tabel 3 menampilkan statistik deskriptif variabel utama. Harga penutupan memiliki fluktuasi moderat, sedangkan volume menunjukkan variabilitas sangat tinggi, dengan standar deviasi besar. Perbedaan karakteristik ini memperkuat perlunya penggunaan variabel turunan untuk menangkap dinamika perubahan, volatilitas, dan tren, sehingga fitur yang dihasilkan lebih informatif untuk model prediksi [10], [25].



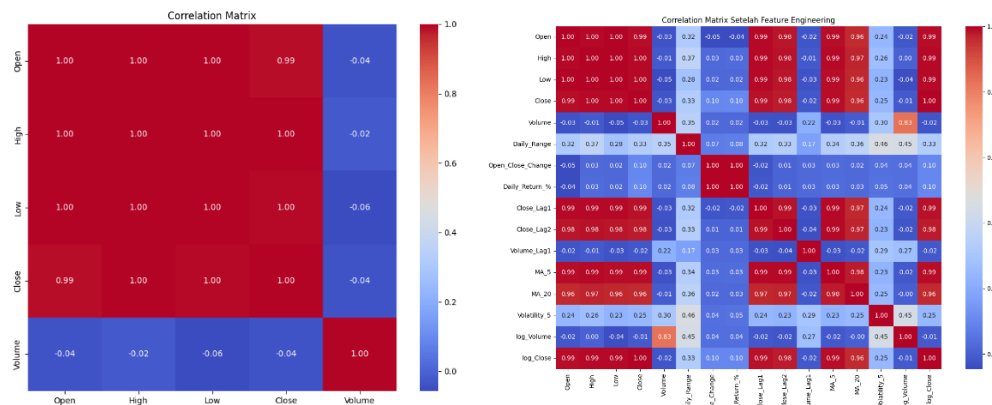
Gambar 2. (Time series) harga penutupan (Close) dan volume



Gambar 3. Distribusi Close & Log-Close

Tabel 3. Statistik Deskriptif

Variabel	Mean	Std Dev	Min	Max
Close	8.961	0.733	7.412	10.570
Open	8.964	0.731	7.412	10.522
High	9.031	0.739	7.504	10.570
Low	8.896	0.721	7.366	10.401
Volume	78.13M	55.48M	10.65M	756.43M



Gambar 4. Correlation Matrix

Gambar 4 menggambarkan correlation matrix antara variabel asli dan variabel turunan hasil feature engineering. Variabel harga (Open, High, Low, Close) memiliki korelasi sangat tinggi ($>0,95$), yang menunjukkan potensi multikolinearitas. Kondisi ini menegaskan perlunya pendekatan regresi dengan regularisasi seperti Ridge dan Lasso, yang telah terbukti efektif mengatasi multikolinearitas pada data finansial [6], [21].

Sementara itu, variabel turunan seperti Daily Range, Daily Return, Moving Average (MA5, MA20), dan Volatility_5 menunjukkan pola korelasi yang lebih beragam dan tidak redundant. Hal ini berarti feature engineering memberikan informasi tambahan yang tidak sepenuhnya ditangkap oleh variabel mentah, sehingga berpotensi meningkatkan performa prediktif model. Temuan ini sejalan dengan literatur yang menekankan bahwa variabel turunan mampu mengurangi noise dan menambah representasi sinyal pasar, terutama pada data saham dengan deret waktu yang kompleks [20], [22].

Secara keseluruhan, hasil EDA menunjukkan bahwa struktur data BBKA memiliki karakteristik umum data keuangan, yaitu tren pada harga, volatilitas pada volume, korelasi tinggi antarvariabel harga, dan pola distribusi yang tidak sepenuhnya normal. Oleh karena itu, transformasi log, pembuatan variabel turunan, dan pemilihan model regularisasi menjadi pendekatan yang tepat dan metodologis kuat, konsisten dengan hasil penelitian terdahulu [11], [21].

3.2 Hasil Evaluasi Model

Untuk menilai kualitas model prediksi saham BBKA, penelitian ini menggunakan tiga metrik evaluasi utama: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Koefisien Determinasi (R^2). Ketiga metrik ini dipilih karena mampu memberikan penilaian yang komprehensif terhadap performa model, baik dari segi akurasi prediksi maupun kemampuan dalam merepresentasikan keragaman data [9], [26]. RMSE dan MAE mengukur besarnya deviasi antara nilai aktual dan prediksi, dengan RMSE lebih sensitif terhadap outlier [21]. Sementara itu, R^2 menunjukkan proporsi variansi data yang berhasil dijelaskan oleh model, sehingga menjadi indikator penting dalam menilai kekuatan prediktif secara keseluruhan [16].

Selain itu, seluruh metrik evaluasi dalam penelitian ini dihitung pada *skala harga asli* (Close). Hal ini dilakukan dengan mengembalikan hasil prediksi dari skala log ke skala asli menggunakan fungsi `expm1()`. Penjelasan memastikan bahwa nilai RMSE dan MAE berada pada satuan harga yang dapat diinterpretasikan secara praktis [4].

Tabel 4. Evaluasi dan Perbandingan Model (Skala Asli)

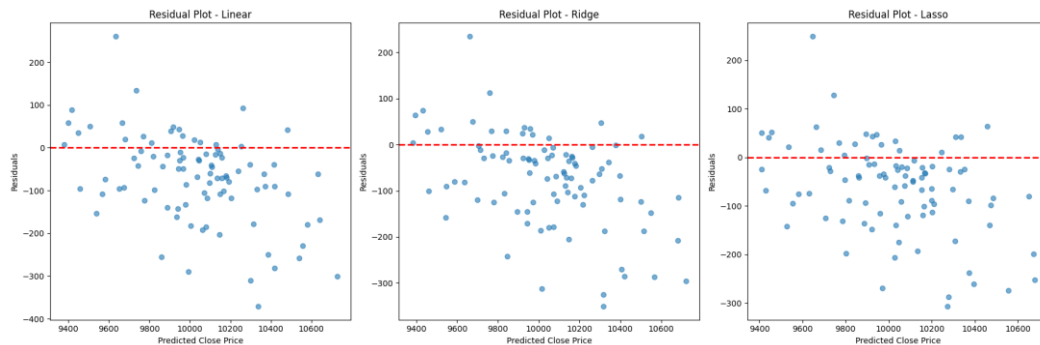
Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
Regresi Linier	67.02	125.53	49.87	94.75	0.9852	0.7850
Ridge Regression	67.75	126.15	50.52	94.87	0.9849	0.7829
Lasso Regression	71.52	113.36	54.20	85.35	0.9832	0.8247

Berdasarkan Tabel 4, Regresi Linier memperoleh performa yang baik pada data latih, tetapi mengalami penurunan signifikan pada data uji sehingga mengindikasikan potensi overfitting [6]. Ridge Regression menunjukkan pola yang hampir sama, sehingga penalti L2 belum memberikan peningkatan berarti. Sebaliknya, Lasso Regression memberikan hasil terbaik pada data uji dengan RMSE lebih rendah dan R^2 tertinggi (0,8247). Temuan ini konsisten dengan karakteristik penalti L1 yang dikenal lebih efektif dalam meningkatkan generalisasi model melalui penyederhanaan struktur koefisien [11].

Hasil tersebut juga selaras dengan EDA yang menunjukkan tingginya korelasi antarvariabel harga. Pada kondisi demikian, model tanpa regularisasi cenderung menghasilkan koefisien yang tidak stabil. Lasso mampu

menekan variabel yang kurang relevan sehingga model menjadi lebih stabil dan adaptif terhadap data baru, sebagaimana juga dilaporkan dalam penelitian terkait prediksi saham di pasar modal Indonesia [7], [17].

Selain evaluasi kuantitatif melalui RMSE, MAE, dan R^2 , analisis juga diperkuat dengan Residual Plot untuk mengamati pola kesalahan prediksi. Residual yang ideal ditandai dengan penyebaran acak di sekitar garis nol tanpa pola sistematis, yang mencerminkan kestabilan dan ketepatan model dalam menangani data [15].



Gambar 5. Residual Plot

Gambar 5 menampilkan residual plot dari ketiga model regresi. Pada Regresi Linier, residual terlihat cukup tersebar dan menunjukkan kecenderungan peningkatan kesalahan pada nilai prediksi yang lebih tinggi. Pola ini mengindikasikan adanya potensi heteroskedastisitas[12]. Ridge Regression menunjukkan pola sebaran residual yang hampir sama, sehingga penalti L2 belum mampu memberikan peningkatan berarti terhadap stabilitas error [8].

Sebaliknya, *Lasso Regression* memperlihatkan distribusi residual yang lebih terkonsentrasi di sekitar garis nol dan tidak menunjukkan pola sistematis. Kondisi ini mengindikasikan bahwa penalti L1 bekerja efektif dalam menyederhanakan model serta meningkatkan kemampuan generalisasi [11]. Selain pola visual yang lebih rapi, residual Lasso juga tampak lebih stabil pada periode volatilitas tinggi, sehingga mengurangi indikasi heteroskedastisitas dibandingkan Regresi Linier maupun Ridge. Hal ini sejalan dengan karakteristik saham BBKA yang cenderung mengalami fluktuasi volatilitas pada waktu-waktu tertentu [27].

Meskipun uji formal seperti Breusch–Pagan atau Durbin–Watson tidak menjadi keharusan dalam konteks model prediksi berbasis machine learning, interpretasi pola residual tetap penting untuk memastikan bahwa model tidak terpengaruh oleh autokorelasi atau variasi residual yang tidak konstan [3]. Analisis ini memperkuat kesimpulan bahwa Lasso memberikan tingkat stabilitas prediksi yang lebih baik dibandingkan dua model lainnya.

Sebagai upaya untuk memperjelas pengaruh regularisasi L1 terhadap struktur model, penelitian ini menyertakan tabel koefisien Lasso Regression. Penyajian koefisien ini memberikan gambaran mengenai fitur yang tetap berkontribusi dalam model serta fitur yang koefisiennya ditekan menuju nol sebagai akibat proses regularisasi [11], [18].

Tabel 5. Seleksi Fitur Lasso

Fitur	Koefisien	Terpilih
Volume	0.000000	Dihilangkan
Daily_Range	0.000000	Dihilangkan
Open_Close_Change	0.000000	Dihilangkan
Daily_Return_%	0.007163	Dipakai
Close_Lag1	0.043356	Dipakai
Close_Lag2	0.000000	Dihilangkan
Volume_Lag1	0.000052	Dipakai
MA_5	0.017596	Dipakai
MA_20	0.000000	Dihilangkan
Volatility_5	0.000000	Dihilangkan
log_Volume	0.000000	Dihilangkan

Tabel 5 tersebut menunjukkan bahwa penalti L1 pada Lasso Regression secara sistematis menekan nilai beberapa koefisien hingga mencapai nol, seperti pada variabel *Volume*, *Daily_Range*, *Close_Lag2*, *MA_20*, dan *Volatility_5*. Kondisi ini mencerminkan mekanisme seleksi fitur yang menjadi karakteristik utama Lasso, di mana variabel yang kontribusinya lemah terhadap model dihilangkan secara otomatis [18]. Sebaliknya, fitur yang tetap mempertahankan koefisien positif, seperti *Close_Lag1*, *Daily_Return_%*, dan *MA_5*, mengindikasikan bahwa variabel tersebut memiliki relevansi yang lebih kuat dalam menjelaskan dinamika harga penutupan BBKA.

Proses regularisasi ini menghasilkan model yang lebih ringkas, terhindar dari redundansi variabel, namun tetap mampu mempertahankan tingkat akurasi yang baik [2]. Dengan demikian, analisis residual ini memperkuat temuan kuantitatif sebelumnya, bahwa Lasso Regression merupakan model dengan performa paling stabil dan akurat dalam memprediksi harga penutupan saham BBKA dibandingkan dengan Linear maupun Ridge Regression [28].

3.3 Hasil Implementasi dengan Streamlit

Implementasi model prediktif dilakukan melalui aplikasi interaktif berbasis Streamlit yang berfungsi sebagai antarmuka terpadu untuk menjalankan keseluruhan proses prediksi secara end-to-end. Aplikasi ini tidak hanya memuat model regresi terbaik, tetapi juga mengintegrasikan seluruh pipeline praproses yang digunakan pada tahap pelatihan, sehingga mekanisme inferensi berjalan konsisten dengan prosedur pemodelan [8], [15]. Gambar 6 memperlihatkan tampilan utama aplikasi yang terdiri atas komponen input, panel prediksi, dan visualisasi grafik.

Agar aplikasi berfungsi secara konsisten dengan proses pelatihan model, alur kerja Streamlit dirancang mengikuti tahapan teknis terstruktur sebagai berikut.

1. Pemuatan Model dan Pipeline

Pada saat aplikasi dijalankan, *Streamlit* melakukan inisialisasi dengan memuat berkas `.joblib` yang berisi model Lasso Regression beserta pipeline lengkapnya. Pipeline tersebut mencakup transformasi logaritmik, standarisasi fitur, penyesuaian urutan variabel, dan prosedur praproses lainnya yang digunakan selama proses pelatihan. Integrasi ini memastikan bahwa seluruh tahapan pengolahan data, mulai dari transformasi hingga scaling, dilakukan secara otomatis dan konsisten [5], [7].

2. Input Fitur oleh Pengguna

Aplikasi menyediakan formulir untuk memasukkan seluruh fitur hasil rekayasa (feature engineering), seperti *Daily Range*, *Open–Close Change*, *Daily Return*, fitur lag, *moving average*, volatilitas rolling window, serta fitur logaritmik. Input tersebut ditampilkan kembali dalam bentuk tabel untuk menjamin kesesuaian struktur data dengan kebutuhan pipeline [26], [28].

3. Penyusunan Data dan Praproses Otomatis

Setelah pengguna memberikan input, sistem menyusun `DataFrame` dengan urutan fitur yang sesuai dan menjalankan seluruh tahapan praproses yang telah dipelajari model. Dengan demikian, data yang masuk diproses dengan cara yang identik dengan data latih, sehingga prediksi yang dihasilkan tetap valid dan tidak mengalami bias akibat ketidaksesuaian transformasi [5], [7].

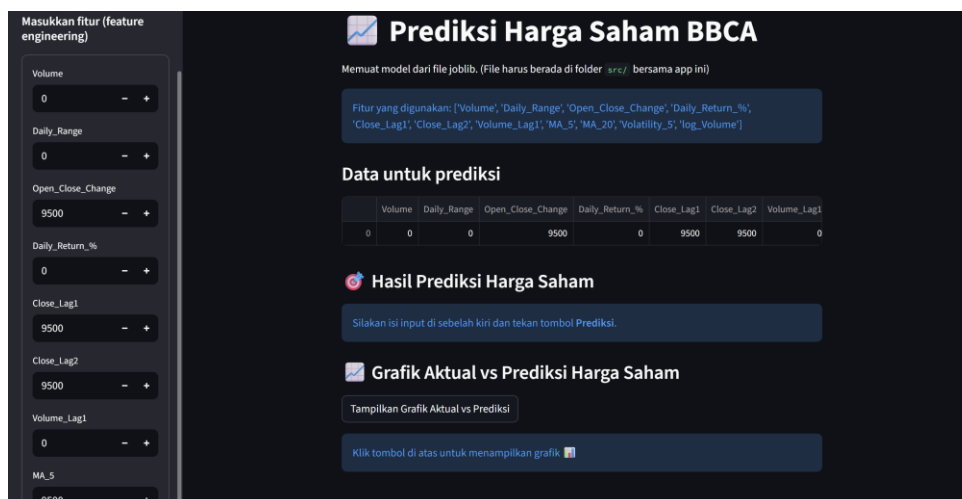
4. Proses Inferensi Model

Ketika pengguna menekan tombol prediksi, data yang telah dipraproses dikirimkan ke model untuk dilakukan inferensi. Model kemudian menghasilkan estimasi harga penutupan saham BBKA. Hasil prediksi ditampilkan pada panel khusus yang memudahkan pengguna dalam meninjau output secara langsung [10], [25].

5. Visualisasi Interaktif

Aplikasi juga menyediakan visualisasi interaktif berupa grafik perbandingan antara harga aktual dan harga hasil prediksi. Grafik ini memungkinkan pengguna untuk memahami pola prediksi dan konsistensi model terhadap data historis melalui pendekatan visual yang informatif dan mudah diinterpretasikan [8], [28].

Dengan demikian, aplikasi Streamlit yang dikembangkan dalam penelitian ini berfungsi sebagai platform prediksi yang terintegrasi, mulai dari pemuatan model dan pipeline, penyediaan input fitur, pelaksanaan praproses otomatis, hingga proses inferensi dan visualisasi hasil. Integrasi tersebut memastikan bahwa model tidak hanya menghasilkan prediksi yang akurat, tetapi juga dapat digunakan secara efisien dan mudah dipahami oleh pengguna akhir.



Gambar 6. Visualisasi Streamlit

4. KESIMPULAN

Penelitian ini dilakukan untuk menjawab kebutuhan model prediksi harga penutupan saham BBKA yang akurat dan stabil, khususnya pada data dengan volatilitas tinggi dan korelasi kuat antarvariabel. Sejalan dengan tujuan penelitian, tiga pendekatan regresi—Regresi Linier, Ridge, dan Lasso—dibandingkan berdasarkan performa prediksi pada data tahun 2023–2024. Hasil analisis menunjukkan bahwa Lasso Regression memberikan performa terbaik, dengan nilai RMSE dan MAE paling rendah serta R^2 tertinggi pada data uji. Keunggulan ini disebabkan oleh mekanisme penalti L1 yang mampu melakukan seleksi variabel secara otomatis, sehingga model menjadi lebih efisien dan memiliki kemampuan generalisasi yang lebih baik dibandingkan Regresi Linier maupun Ridge. Secara ilmiah, penelitian ini memberikan kontribusi berupa bukti empiris tentang efektivitas regularisasi L1 dalam prediksi harga saham di pasar modal Indonesia, terutama pada dataset yang memiliki multikolinearitas tinggi. Selain itu, penelitian ini menghasilkan aplikasi Streamlit yang mengintegrasikan seluruh model regresi beserta pipeline praprosesnya, sehingga memungkinkan pengguna melakukan input fitur, memperoleh prediksi, dan meninjau visualisasi hasil secara interaktif. Integrasi ini memperkuat aspek implementatif penelitian dan meningkatkan aksesibilitas model bagi pengguna non-teknis. Keterbatasan utama penelitian terletak pada cakupan data yang relatif singkat dan variabel yang masih terbatas pada harga historis serta volume. Untuk penelitian selanjutnya, disarankan memperluas rentang waktu data, menambahkan indikator eksternal seperti variabel makroekonomi dan sentimen pasar, serta mengeksplorasi metode prediksi non-linier atau hybrid. Pengembangan tersebut diharapkan dapat menghasilkan model yang lebih akurat, robust, dan adaptif terhadap dinamika pasar modal Indonesia yang kompleks.

REFERENCES

- [1] E. Siska, O. Duraipandi, and P. Widodo, "Determinants of Indonesian stock market development: Implementation of an ARDL bound testing approach," *Investment Management and Financial Innovations*, vol. 20, no. 4, pp. 69–82, Oct. 2023, doi: 10.21511/imfi.20(4).2023.07.
- [2] T. B. Shahi, A. Shrestha, A. Neupane, and W. Guo, "Stock price forecasting with deep learning: A comparative study," *Mathematics*, vol. 8, no. 9, p. 1441, Aug. 2020, doi: 10.3390/math8091441.
- [3] S. A. L. Satriyo, A. R. Pratama, and Rahmat, "Perbandingan metode linear regresi dan polynomial regresi untuk memprediksi harga saham studi kasus Bank BCA," *INFOTECH: Journal of Information Technology*, vol. 4, no. 1, Jun. 2023, doi: 10.37373/infotech.v4i1.602.
- [4] S. Soewignjo, Sediono, M. F. F. Mardianto, and E. Pusporani, "Prediksi harga saham Bank BCA (BBKA) pasca stock split dengan artificial neural network dengan algoritma backpropagation," *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 4, pp. 1683–1693, Oct. 2023, doi: 10.33379/gtech.v7i4.3363.
- [5] E. M. Raouhi, M. Lachgar, and A. Kartit, "Comparative study of regression and regularization methods: Application to weather and climate data," in *WITS 2020*, S. Bennani et al., Eds. Singapore: Springer, 2022, pp. 233–240, doi: 10.1007/978-981-33-6893-4_22.
- [6] S. K. Safi et al., "Optimizing linear regression models with Lasso and Ridge regression: A study on UAE financial behavior during COVID-19," *Migration Letters*, vol. 20, no. 6, pp. 139–153, Sep. 2023, doi: 10.59670/ml.v20i6.3468.
- [7] A. J. Daniel, M. D. Ibrahim, and A. A. Josaphat, "Modeling stock data using multiple linear regression and LASSO regression analysis," *Mikailalsys Journal of Mathematics and Statistics*, vol. 3, no. 2, pp. 490–499, Jun. 2025, doi: 10.58578/mjms.v3i2.5927.
- [8] E. A. Gultom, K. D. S. Susilowati, and A. Kusmintarti, "Design of a stock forecasting dashboard using Python-Streamlit and FB Prophet with AI," *Formosa Journal of Science and Technology*, vol. 3, no. 11, pp. 2445–2464, Nov. 2024, doi: 10.55927/fjst.v3i11.12216.
- [9] P. Mittal, S. Kanimozhi, and L. Sairamesh, *Interactive and Dynamic Stock Market Dashboard*. CRC Press, 2024.
- [10] R. Zapar et al., "Penerapan model regresi linier untuk prediksi harga saham Bank BCA pada Bursa Efek Indonesia," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, Feb. 2024, doi: 10.36040/jati.v8i1.8215.
- [11] S. D. Saputra and A. D. Widianoro, "BBKA stock price prediction using linear regression method," *International Journal of Artificial Intelligence Science*, vol. 1, no. 1, pp. 25–36, Sep. 2024, doi: 10.63158/IJAIS.v1.i1.7.
- [12] R. Anbu, R. Krishnamoorthi, and A. KamalNath, "A comparative analysis of linear regression and LSTM for stock price prediction," *International Journal of Research Publication and Reviews*, vol. 6, no. 5, pp. 15122–15128, 2025.
- [13] M. Q. Shobri et al., "Model analisis harga saham sektor finansial PT Bank Central Asia Tbk (BBKA)," *Jurnal Keilmuan dan Keislaman*, pp. 260–269, Dec. 2023, doi: 10.23917/jkk.v2i4.170.
- [14] J. Iworiso, "Forecasting stock market out-of-sample with regularised regression training techniques," *International Journal of Economics, Finance and Management*, vol. 11, no. 1, pp. 1–12, May 2023, doi: 10.12691/ijefm-11-1-1.
- [15] R. Buga et al., "Streamlit application and deep learning model for brain metastasis monitoring after gamma knife treatment," *Biomedicine*, vol. 13, no. 2, p. 423, Feb. 2025, doi: 10.3390/biomedicine13020423.
- [16] D. Ekambaram and V. Ponnusamy, "Real-time monitoring and assessment of rehabilitation exercises for low back pain through interactive dashboard pose analysis using Streamlit—A pilot study," *Electronics*, vol. 13, no. 18, p. 3782, Sep. 2024, doi: 10.3390/electronics13183782.
- [17] B. Iraqi, L. Benhiba, and M. A. J. Idrissi, "Data analytics in investment banks," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021, doi: 10.14569/IJACSA.2021.0120562.
- [18] E. C. Wibowo and A. D. Cahyono, "Analisis perbandingan algoritma regresi linier dengan neural network untuk prediksi harga saham," *Jurnal Sistem Informasi*, vol. 14, no. 4, pp. 1879–1896, 2025.

- [19] X. Li et al., "Evaluation of supervised machine learning regression models for CFD-based surrogate modelling in indoor airflow field reconstruction," *Building and Environment*, vol. 267, p. 112173, Jan. 2025, doi: 10.1016/j.buildenv.2024.112173.
- [20] X. Wang, W. Wang, and S. Zhang, "Stock price return prediction based on multifactorial machine learning approaches," in *ICBBEM 2022*, Atlantis Press, 2023, pp. 324–333, doi: 10.2991/978-94-6463-030-5_34.
- [21] C. N. C. Neba et al., "Comparative analysis of stock price prediction models: GLM, Ridge, Lasso, ElasticNet, and Random Forest—Case study on Netflix," Oct. 2023, doi: 10.5281/zenodo.10040460.
- [22] F. F. Abdulloh et al., "Performance analysis of machine learning algorithms using the ensemble method on predicting the impact of inflation on Indonesia's economic growth," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 4, p. 2316, Dec. 2024, doi: 10.62527/joiv.8.4.2567.
- [23] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.
- [24] K. Abiodun et al., "Risk-sensitive financial dashboards with embedded machine learning: A user-centric approach to operational transparency," *International Journal of Scientific Research in Modern Technology*, pp. 1–18, Feb. 2024, doi: 10.38124/ijsrmt.v3i2.678.
- [25] Y. Anisa, M. Hafiz, and N. Novita, "Pengembangan model prediksi harga saham dengan menggunakan regresi linear berganda pada saham BRI," *JISTech: Journal of Islamic Science and Technology*, vol. 9, no. 2, p. 191, Dec. 2024, doi: 10.30829/jistech.v9i2.22213.
- [26] C. K. Bal and R. K. Mishra, "Stock market analysis and prediction: A bibliometric analysis," *Journal of Scientometric Research*, vol. 14, no. 1, pp. 221–238, Mar. 2025, doi: 10.5530/jscires.20251277.
- [27] L. Pratt, C. Bisson, and T. Warin, "Bringing advanced technology to strategic decision-making: The DI/DS integration framework," *Futures*, vol. 152, p. 103217, Sep. 2023, doi: 10.1016/j.futures.2023.103217.
- [28] M. Du, R. Amor, K.-L. Ma, and B. C. Wünsche, "Data visualization for improving financial literacy: A systematic review," *Visual Informatics*, p. 100272, Sep. 2025, doi: 10.1016/j.visinf.2025.100272.