



# Implementasi Regresi Linier Berganda Untuk Prediksi Harga Mobil Bekas Di Indonesia Berbasis Gradio

M Ridho Alfani, Elvia Budianita\*, Lestari Handayani, Siti Ramadhani

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1</sup>12150113913@students.uin-suska.ac.id, <sup>2,\*</sup>elvia.budianita@uin-suska.ac.id, <sup>3</sup>lestari.handayani@uin-suska.ac.id,

<sup>4</sup>siti.ramadhani@uin-suska.ac.id

Email Penulis Korespondensi: elvia.budianita@uin-suska.ac.id

**Abstrak**—Harga kendaraan bekas bergantung pada berbagai aspek yang menyebabkan perubahan nilai jual di pasar, seperti model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc. Permasalahan yang sering terjadi dalam transaksi mobil bekas adalah menentukan harga yang masih belum sepenuhnya berbasis analisis data yang terukur. Tujuan dari penelitian ini adalah merancang model untuk memperkirakan harga mobil bekas melalui metode regresi linier berganda serta mengimplementasikannya dalam User Interface. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari repositori publik Kaggle, serta dikumpulkan dari beberapa forum jual beli mobil bekas di Pekanbaru dan platform media sosial seperti Facebook yang memuat informasi harga kendaraan. Dataset tersebut berisi 400 baris data dengan rentang tahun kendaraan mobil dari 2005 hingga 2025. Tahapan penelitian meliputi praproses data berupa encoding variabel kategorikal dan normalisasi data, Data dibagi menjadi data pelatihan dan data pengujian, dilanjutkan dengan proses pembentukan model serta penilaian performa model. Evaluasi dilakukan dengan menggunakan metrik Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi ( $R^2$ ). Model dibangun menggunakan beberapa variabel independen, yaitu model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc, dengan harga kendaraan sebagai variabel dependen. Berdasarkan hasil pengujian, metode regresi linier berganda menunjukkan kemampuan untuk menghasilkan estimasi harga mobil bekas serta memiliki potensi penerapan dalam sistem pendukung keputusan. Hasil pengujian menunjukkan oleh nilai MSE pada data training sebesar 0,004 dan data testing sebesar 0,010, MAE pada data latih sebesar 0,046 dan data uji sebesar 0,071, dan RMSE pada data latih sebesar 0,062 dan sebesar 0,100 pada data uji, koefisien determinasi ( $R^2$ ) pada data latih sebesar 0,985 dan pada data uji sebesar 0,955. Model selanjutnya diimplementasikan menggunakan library Python Gradio agar pengguna dapat melakukan prediksi harga kendaraan melalui User Interface.

**Kata Kunci:** Gradio; Mobil Bekas; MSE; Mean Squared Error; Prediksi Harga; Regresi Linier Berganda

**Abstract**—The price of a used vehicle depends on various aspects that cause changes in the selling value in the market, such as model, year, transmission, mileage, fuel, tax, mpg, and cc. A common problem in used car transactions is determining prices that are still not fully based on measurable data analysis. The purpose of this study is to design a model to estimate the price of a used car through the multiple linear regression method and implement it in the User Interface. The data used in this study is secondary data obtained from the Kaggle public repository, and collected from several used car buying and selling forums in Pekanbaru and social media platforms such as Facebook that contain vehicle price information. The dataset contains 400 rows of data with a range of car years from 2005 to 2025. The research stages include data preprocessing in the form of categorical variable encoding and data normalization. Data is divided into training data and testing data, followed by the process of model building and model performance assessment. Evaluation is carried out using the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination ( $R^2$ ) metrics. The model was built using several independent variables, namely model, year, transmission, kilometer, fuel, tax, mpg, and cc, with vehicle price as the dependent variable. Based on the test results, the multiple linear regression method shows the ability to produce used car price estimates and has potential for application in decision support systems. The test results show that the MSE value on the training data is 0.004 and the testing data is 0.010, MAE on the training data is 0.046 and the testing data is 0.071, and RMSE on the training data is 0.062 and 0.100 on the testing data, the coefficient of determination ( $R^2$ ) on the training data is 0.985 and on the testing data is 0.955. The next model is implemented using the Python Gradio library so that users can predict vehicle prices through the User Interface.

**Keywords:** Gradio; Used Car; MSE; Mean Squared Error; Price Prediction; Multiple Linear Regression

## 1. PENDAHULUAN

Kendaraan mobil saat ini telah menjadi salah satu alat transportasi darat yang paling populer dengan tingkat pemanfaatan yang sangat tinggi bagi masyarakat di Indonesia. Peningkatan kebutuhan mobilitas masyarakat mendorong tingginya permintaan terhadap kendaraan pribadi, khususnya mobil. Kemudahan proses kepemilikan serta kemudahan penggunaan menjadikan mobil sebagai pilihan utama dalam menunjang aktivitas sehari-hari. Selain mobil baru, mobil bekas juga menjadi alternatif utama yang banyak diminati karena menawarkan harga terjangkau, usia kendaraan layak pakai, serta beban pajak yang lebih ringan [1]. Di Indonesia, Kendaraan merupakan sarana transportasi yang memiliki peran sangat penting dan telah menjadi kebutuhan pokok masyarakat dalam mendukung mobilitas sehari-hari bahkan sudah dianggap sebagai kebutuhan primer pada era modern saat ini [2].

Toyota sebagai produsen global yang telah berperan penting dalam kemajuan industri otomotif serta penyediaan unit berkualitas tinggi di pasar nasional [3]. Seiring dengan perkembangan ekonomi dan perubahan gaya hidup masyarakat, kebutuhan terhadap kendaraan pribadi yang andal terus mengalami peningkatan. Kondisi ini menyebabkan pasar mobil bekas menjadi semakin berdaya saing, kompetitif, dan dinamis dari tahun ke tahun. Banyaknya merek dan tipe kendaraan yang tersedia di pasaran, seperti Toyota, Mitsubishi, dan Suzuki, turut meningkatkan tingkat kesulitan bagi pelaku usaha maupun konsumen awam dalam menentukan harga jual kendaraan bekas secara adil dan akurat.



Penentuan harga mobil bekas sangat bergantung pada sejumlah faktor yang kompleks, termasuk model, tahun, transmisi, kilometer, bahan bakar, pajak, mil per gallon (*mpg*), kapasitas mesin (*cc*), yang dikenakan [4].

Pertumbuhan ekonomi masyarakat mendorong semakin banyaknya aktivitas bisnis yang dilakukan untuk meningkatkan kesejahteraan rumah tangga. Seiring perkembangan tersebut, kebutuhan transportasi terus bertambah setiap tahunnya [5]. Permasalahan mendasar yang sering terjadi dalam transaksi mobil bekas saat ini adalah proses menentukan harga yang belum sepenuhnya berbasis pada analisis data yang terukur. Bidang analisis statistik, metode regresi linier sederhana umumnya digunakan untuk menganalisis hubungan antara satu variabel independen dengan satu variabel dependen. Namun, metode ini memiliki keterbatasan karena hanya mampu mempertimbangkan satu faktor tunggal dalam pemodelan, untuk mengatasi keterbatasan tersebut, metode regresi linier berganda dimanfaatkan dalam penelitian ini untuk menganalisis hubungan dari berbagai variabel independen dengan variabel dependen, sehingga mampu menghasilkan nilai estimasi prediksi yang jauh lebih lengkap, rasional, dan komprehensif [6].

Beberapa penelitian terdahulu telah membahas perkiraan harga mobil bekas menggunakan ragam pendekatan metode regresi sebagai instrumen pembanding. Penelitian ini [7] mengembangkan aplikasi prediksi harga mobil bekas berbasis web menggunakan regresi linier dan memperoleh nilai koefisien determinasi ( $R^2$ ) sebesar 0,94 yang menunjukkan tingkat akurasi tinggi. Namun, model pembanding tersebut masih terbatas karena belum menguji variasi data skunder yang dikombinasikan dengan dinamika pasar lokal. Penelitian lain [8] mencoba mengeksplorasi model pembanding menggunakan dataset *Used Car Listings in Indonesia* dengan melibatkan 19 atribut dan 12.500 baris data, namun riset tersebut hanya menghasilkan nilai R-squared sebesar 51% sehingga model belum mampu menjelaskan variasi harga secara optimal.

Sementara itu, penelitian [1] mengindikasikan bahwa regresi linier berganda sangat mampu digunakan untuk melakukan prediksi terhadap tren penjualan dengan akurasi yang memadai, meskipun diakui masih terdapat keterbatasan fungsional dalam menangkap tingkat kesulitan pasar. Riset lainnya, penelitian [9] membahas faktor-faktor yang memengaruhi keputusan pembelian mobil menggunakan hasil pengujian regresi linier berganda membuktikan secara empiris bahwa variabel-variabel yang diteliti memiliki pengaruh yang bermakna secara statistik terhadap variabel dependen keputusan pembelian.

Berdasarkan rangkaian penelitian-penelitian terdahulu tersebut, terdapat kesenjangan yang nyata dalam pemilihan kombinasi variabel independen yang memiliki hubungan kuat terhadap harga serta minimnya implementasi sistem yang mudah diakses oleh masyarakat. Beberapa penelitian pembanding masih menunjukkan tingkat akurasi yang belum optimal ketika dihadapkan pada variasi data lokal, atau belum menerapkan model ke dalam bentuk antarmuka pengguna yang interaktif. Oleh karena itu, pengembangan model estimasi harga pasar mobil bekas dalam penelitian ini bertumpu pada pendekatan regresi linier berganda untuk memetakan keterkaitan terukur antar variabel independen yang digunakan. Kinerja model dalam memproyeksikan nilai kendaraan dianalisis secara ketat menggunakan serangkaian parameter evaluasi, di antaranya  $R^2$  untuk mengukur persentase variansi, serta MAE, RMSE, MSE, untuk meninjau besaran tingkat kesalahan estimasi model pada data latih dan data uji. selain itu, model diintegrasikan secara langsung ke dalam antarmuka interaktif menggunakan library Python Gradio agar pengguna dapat melakukan estimasi harga secara langsung, instan, dan objektif.

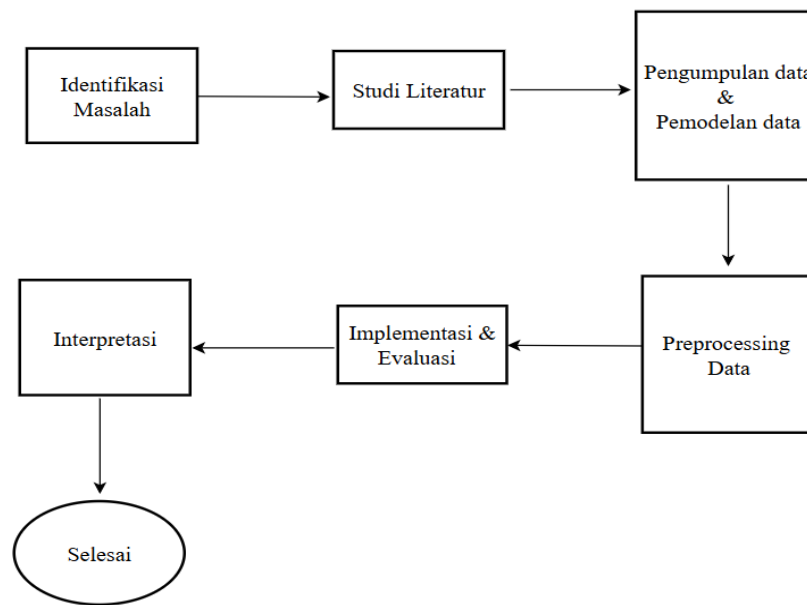
Berdasarkan seluruh uraian konteks permasalahan serta evaluasi terhadap berbagai penelitian terdahulu di atas, maka penelitian ini dilakukan dengan tujuan utama untuk merancang dan membangun model prediksi berbasis metode Regresi Linier Berganda yang mampu memperkirakan nilai harga jual mobil bekas merek Toyota secara, terukur, dan objektif berdasarkan delapan variabel independen penentu (model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc). Melalui analisis statistik ini, penelitian juga bertujuan untuk menganalisis besaran kontribusi statistik dari masing-masing variabel tersebut terhadap perubahan harga, serta melakukan pengujian performa model secara ketat menggunakan metrik validasi standar seperti MSE, RMSE, MAE, dan koefisien determinasi ( $R^2$ ) guna memastikan model terhindar dari kekeliruan prediksi.

Selain itu, tujuan akhir dari riset ini adalah mentransformasikan model tersebut ke dalam bentuk antarmuka pengguna (User Interface) berbasis web yang interaktif menggunakan library Gradio. Hal ini ditujukan untuk menyediakan instrumen pendukung keputusan penentuan harga yang mudah dioperasikan oleh masyarakat luas maupun pelaku bisnis otomotif lokal, sehingga dapat memberikan prediksi nyata dalam meminimalisir risiko kerugian finansial akibat ketidakseimbangan informasi di pasar mobil bekas Indonesia. sekaligus memberikan prediksi harga yang objektif bagi pasar lokal dan referensi akademis bagi penerapan model regresi linear.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan deskriptif, yang diterapkan untuk menganalisis dan mengolah data numerik guna menggambarkan kondisi yang ada dan menjelaskan fenomena yang diteliti untuk mengidentifikasi hubungan serta hubungan variabel independen terhadap variabel dependen. Dalam penelitian ini, variabel independen yang digunakan meliputi model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc, sedangkan variabel dependen yang dianalisis adalah harga jual mobil bekas. Alur metodologi penelitian digambarkan pada Gambar 1.



**Gambar 1.** Metodologi Penelitian Regresi Linier

### 2.1.1 Identifikasi Masalah

Perkembangan bisnis otomotif saat ini membuat persaingan antar perusahaan menjadi semakin ketat. Kondisi ini mendorong pabrikan untuk terus memproduksi mobil-mobil model terbaru demi menarik minat pasar. Selain membantu mobilitas dan perekonomian masyarakat, banyaknya produksi mobil baru ini ternyata ikut membuat jumlah ketersediaan mobil bekas di pasaran meningkat drastis. Masalahnya, harga mobil bekas selalu mengalami penurunan nilai (penyusutan) secara berkala seiring bertambahnya umur mobil tersebut. Ditambah lagi, pembeli sering kesulitan menentukan harga yang pas karena harus mempertimbangkan banyak hal teknis seperti tipe, tahun, transmisi, hingga jarak tempuh kendaraan. Akibatnya, masyarakat sering kali bingung dan ragu dalam mengambil keputusan saat membeli mobil bekas karena takut salah pilih atau mendapat harga yang terlalu mahal. Oleh karena itu, penelitian ini mencoba menerapkan analisis data untuk membuat model prediksi yang bisa membantu masyarakat memperkirakan harga pasar mobil bekas secara lebih objektif [1],[2],[8].

Selain itu, Implementasi regresi linier dalam sistem berbasis web maupun framework aplikasi juga menunjukkan bahwa metode ini memiliki keunggulan dari sisi kesederhanaan, efisiensi komputasi, dan kemudahan interpretasi hasil, sehingga tetap relevan diterapkan guna merancang skema estimasi nilai jual kendaraan [7]. Permasalahan utama diangkat pada studi ini berfokus terhadap bagaimana menerapkan pendekatan regresi linier berganda guna memprediksi harga jual mobil bekas berdasarkan atribut kendaraan secara akurat, serta mengukur performa model dengan metrik MSE.

### 2.1.2 Studi Literatur

Penelitian terdahulu mengindikasikan bahwa metode regresi linier berganda mampu dimanfaatkan secara optimal dalam melakukan prediksi harga mobil bekas berdasarkan berbagai variabel penentu. Variabel seperti tahun dan cc diketahui memiliki pengaruh yang signifikan terhadap harga kendaraan. Selain itu, metode ini mampu menggambarkan hubungan antara variabel independen dan harga secara jelas. Dalam konteks industri otomotif Indonesia, regresi linier berganda dinilai relevan untuk menganalisis pola dan tren harga kendaraan [3],[5],[10].

Dalam konteks pengembangan sistem berbasis machine learning, beberapa penelitian membandingkan model regresi linier dengan metode ensemble learning dan menunjukkan bahwa regresi linier tetap unggul dari sisi kejelasan interpretasi. Selain itu, penerapan regresi linier berganda pada berbagai bidang, seperti prediksi CPO dan inflasi, terbukti menghasilkan performa yang stabil dan dapat diandalkan. Penelitian lain juga menegaskan kemampuan beradaptasi regresi linier berganda dalam memodelkan berbagai jenis data numerik dengan hubungan yang jelas antarvariabel [11],[12],[13].

Berdasarkan hasil studi literatur tersebut, dapat disimpulkan bahwa regresi linier berganda menjadi pendekatan yang sesuai guna menganalisis hubungan antar variabel yang memengaruhi harga jual mobil bekas. Akan tetapi, sebagian besar studi terdahulu masih memiliki keterbatasan pada jumlah parameter yang digunakan serta belum memanfaatkan data berskala besar seperti dataset publik dari Kaggle dan beberapa sorum mobil bekas. Selain itu, beberapa penelitian sebelumnya belum menggunakan metrik MSE sebagai ukuran performa utama model. Dengan demikian, penelitian ini mengembangkan model regresi linier berganda menggunakan pustaka Scikit-learn di Google Colab dan melakukan evaluasi model menggunakan MSE untuk menghasilkan prediksi harga mobil bekas yang lebih objektif dan terukur. Setelah menentukan metode dan sumber data penelitian, pada tahap selanjutnya, dilakukan pengumpulan data yang menjadi dasar dalam proses pemodelan serta evaluasi model prediksi.



### 2.1.3 Pengumpulan Data

Sumber informasi pada riset ini merupakan data sekunder yang diambil melalui repositori publik Kaggle <https://share.google/6fGMMsU51K92FWsdK>. Selain itu, data juga dikumpulkan dari beberapa forum jual beli mobil bekas di Pekanbaru serta platform media sosial seperti Facebook yang memuat informasi harga kendaraan. Dataset tersebut berisi informasi mengenai harga jual mobil bekas dengan total 400 baris data dan mencakup rentang tahun kendaraan dari 2005 hingga 2025 mobil merek toyota. Dataset tersebut dimanfaatkan sebagai pijakan dalam penyusunan skema prediksi harga kendaraan mengacu pada regresi linier berganda. Rentang tahun yang diperluas hingga tahun 2025 bertujuan untuk meningkatkan relevansi data terhadap kondisi pasar terkini, sehingga model yang dibangun mampu menangkap pola harga kendaraan yang lebih baru. Dataset ini selanjutnya digunakan dalam proses pengolahan awal, tahap distribusi dataset ke dalam kelompok training dan testing, diikuti penyusunan model regresi untuk menghasilkan sistem prediksi harga yang akurat.

Struktur dan karakteristik dataset mobil bekas yang digunakan dalam proses pemodelan ini dirangkum pada Tabel 1. Penyajian data tersebut bertujuan untuk memberikan gambaran yang terstruktur mengenai setiap variabel sebelum dilakukan analisis lebih lanjut.

**Tabel 1.** Dataset Mobil

Model	Tahun	Harga	Transmisi	Kilometer	Bahan Bakar	Pajak	Mpg	Cc
Agya	2013	85188129	manual	234612	bensin	2300000	40.2	1
Agya	2013	85250230	matic	212245	bensin	2300001	40.2	1
Innova	2017	242000000	manual	87000	bensin	4900000	1.11	2
Hilux DC	2022	405000000	manual	4578	diesel	6000000	1.11	2.4

Berdasarkan rincian data pada Tabel 1 tersebut, struktur data memuat beberapa atribut penting yang berperan sebagai variabel dalam penelitian, yaitu model kendaraan, tahun, harga, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc. Setiap atribut memiliki peran dalam memengaruhi harga jual kendaraan, di mana harga digunakan sebagai variabel dependen, sedangkan atribut lainnya sebagai variabel independen. Melalui pembagian variabel yang tertera sistem dapat mengenali pola matematis dari kombinasi tipe transmisi, konsumsi bahan bakar (mpg), hingga kapasitas mesin (cc) untuk menghasilkan estimasi harga prediksi.

### 2.1.4 Preprocessing Data

Tahap langkah ini dimaksudkan guna memastikan integritas dataset serta layak dianalisis. Adapun langkah-langkah yang dilakukan meliputi :

a. **Membersihkan Data (Data Cleaning):**

Tahap pembersihan data dalam preprocessing dilakukan untuk mengidentifikasi serta menangani data yang tidak valid, tidak lengkap, serta data yang tidak relevan sehingga kualitas data menjadi lebih baik untuk dianalisis [12].

b. **Transformasi Data:**

Mengubah atribut kategorikal seperti transmisi dan bahan bakar menjadi bentuk numerik menggunakan metode One Hot Encoding, sehingga setiap kategori direpresentasikan dalam bentuk variabel biner (0 dan 1) untuk menghindari asumsi hubungan ordinal antar kategori [14].

c. **Normalisasi Data :**

Menyamakan skala nilai antar variabel numerik agar model dapat di pelajari hubungan antarvariabel dengan lebih baik.

d. **Pembagian Data ( Split Data ):**

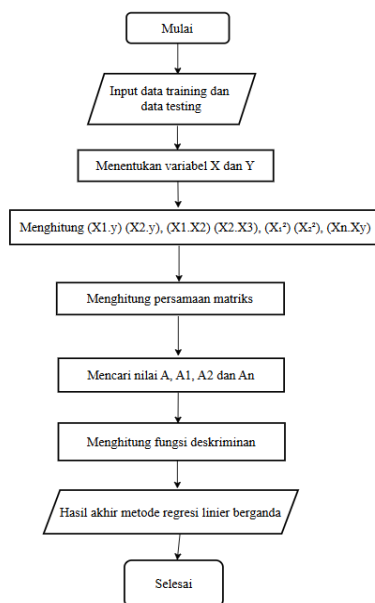
Pada tahap prapemrosesan, kumpulan data didistribusikan ke dalam sampel pelatihan serta pengujian. Sampel pelatihan difungsikan guna membangun kerangka regresi, sementara data pengujian berperan dalam memvalidasi performa skema terhadap data yang belum pernah dilihat sebelumnya.

### 2.1.5 Seleksi Variabel

Penelitian ini melibatkan seluruh variabel independen yang tersedia tanpa melalui proses eliminasi atau seleksi fitur. Kedelapan atribut tersebut meliputi model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc digunakan secara bersamaan karena jumlahnya yang relatif sedikit dan secara teoritis seluruhnya berpengaruh langsung terhadap harga jual mobil bekas Toyota. Kontribusi dari seluruh variabel ini nantinya akan dievaluasi secara bersamaan menggunakan nilai Koefisien Determinasi ( $R^2$ ) pada model regresi.[13].

### 2.1.6 Pemodelan Data

Tahap ini menjadi fokus utama penelitian karena melibatkan implementasi regresi linier berganda memanfaatkan modul Scikit-learn pada lingkungan Google Colab. Sebelum proses pemodelan dilakukan, data terlebih dahulu melalui tahap preprocessing yang mencakup pembersihan data, transformasi nilai kategorikal menjadi numerik agar model dapat belajar secara optimal. Model ini dibangun berdasarkan variabel independen seperti tipe kendaraan, tahun rilis, jenis transmisi, jarak tempuh, tipe bahan bakar, pajak, mpg, dan cc. Flowchart proses pemodelan data ditunjukkan pada Gambar 2 berikut ini.



**Gambar 2.** Flowchart proses regresi linier berganda

Flowchart tersebut menggambarkan tahapan proses pemodelan yang meliputi:

- Mulai : Titik awal proses analisis.
- Input data training dan testing : Menyiapkan data mobil (tahun, kilometer, pajak, dll.)
- Menentukan variabel  $X$  dan  $Y$  :  $X$  = variabel bebas,  $Y$  = variabel harga
- Menghitung  $(X_1 \cdot Y), (X_2 \cdot Y), (X_1 \cdot X_2), (X_2 \cdot X_3), (X_1^2), (X_2^2), (X_n \cdot X_n)$  : Persiapan untuk membentuk matriks regresi
- Menghitung persamaan matriks  $(A)$  : Membangun sistem persamaan regresi
- Mencari nilai  $A, A_1, A_2, \dots, A_n$  : Menghitung koefisien menggunakan  $(\det(A_0)/\det(A), dst.)$
- Menghitung fungsi diskriminan / hasil regresi : Membentuk persamaan akhir  $Y = A + A_1X_1 + A_2X_2 + \dots + A_nX_n$
- Hasil akhir prediksi regresi linier : Menampilkan hasil (misalnya prediksi harga mobil)
- Selesai : Akhir proses

Langkah-langkah tersebut diambil dari metode yang dijelaskan dalam penelitian sebelumnya yang menggunakan tahapan serupa dalam pengembangan model regresi linier berganda untuk analisis data kuantitatif.

Sebagai salah satu metode penting dalam statistika terapan dan komputasi pembelajaran mesin, regresi linear diterapkan untuk mengukur hubungan fungsional antara variabel independen dan target dependen. Hubungan tersebut dinyatakan melalui bentuk persamaan linear, di mana variabel dependen ditempatkan secara tunggal pada salah satu sisi persamaan. Ketika analisis melibatkan dua atau lebih variabel independen ( $X$ ), model ini berkembang menjadi Regresi Linear Berganda. Penerapan model multivariat ini bertujuan untuk memvalidasi hubungan antar-elemen data, sekaligus memproyeksikan sejauh mana fluktuasi kolektif pada variabel prediktor memengaruhi nilai variabel dependen ( $Y$ ) [10], [11].

Dalam kajian ini, model regresi linier berganda diaplikasikan untuk dibangun berdasarkan hubungan antara variabel independen (tahun, kilometer, cc, bahan bakar, dan transmisi) dengan variabel dependen (harga jual mobil bekas). Regresi linier berganda adalah salah satu metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen ( $Y$ ) dengan lebih dari satu variabel independen ( $X_1, X_2, \dots, X_n$ ). Persamaan umum regresi linier berganda adalah:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n + \varepsilon \quad (1)$$

Dalam persamaan tersebut, simbol  $Y$  merepresentasikan variabel dependen yang merujuk pada nilai estimasi harga mobil bekas. Konstanta atau intercept disimbolkan oleh  $\beta_0$ , yang menunjukkan nilai  $Y$  ketika seluruh prediktor bernilai nol. Sementara itu, notasi  $\beta_1$  hingga  $\beta_n$  merupakan koefisien regresi yang mencerminkan besaran pengaruh dari masing-masing variabel independen  $X_1$  sampai  $X_n$ . Karakteristik teknis kendaraan seperti tahun perakitan, jarak tempuh (kilometer), jenis bahan bakar, tipe transmisi, serta kapasitas mesin (cc) bertindak sebagai variabel independen tersebut. Model ini juga dilengkapi dengan komponen  $\varepsilon$  (error residual) untuk menampung variasi data yang tidak mampu dijelaskan oleh variabel bebas.

### 2.1.7 Implementasi

Dalam pelaksanaan penelitian ini memanfaatkan sejumlah alat dan bahan untuk menunjang kegiatan pengolahan data serta pembangunan model prediksi harga mobil bekas. Alat penelitian berfungsi sebagai sarana untuk menjalankan proses analisis data, pembangunan model, serta implementasi sistem prediksi, dalam penelitian ini, peralatan yang digunakan mencakup hardware dan software. Perangkat keras yang dipakai antara lain komputer atau laptop yang dimanfaatkan



dalam proses penelitian untuk menjalankan proses pengolahan data dan implementasi sistem. Sedangkan perangkat lunak yang digunakan meliputi bahasa pemrograman Python sebagai bahasa utama dalam pembangunan model prediksi, Google Colab sebagai lingkungan pengembangan untuk menjalankan kode program, serta beberapa manajemen data dilakukan memanfaatkan Pandas dan NumPy, sedangkan Scikit-learn untuk pembangunan model Regresi Linier Berganda, dan Gradio untuk membangun antarmuka pengguna yang interaktif, bahan penelitian ini berupa dataset harga mobil bekas yang diperoleh dari sumber data publik.

Dataset tersebut memuat sejumlah atribut yang berperan sebagai variabel independen, meliputi model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc. Adapun variabel dependen dalam penelitian ini adalah harga mobil bekas. Penyusunan model prediksi ini mengacu pada dataset yang telah dikumpulkan, tahap implementasi sistem diawali dengan preprocessing data yang termasuk di dalamnya penerapan skema encoding untuk mengubah variabel non-numerik, serta langkah pembersihan dataset, serta partisi dataset ke dalam sampel pelatihan dan kelompok pengujian. Sampel pelatihan dipakai dalam proses pembelajaran model Regresi Linier Berganda untuk mengekstraksi pola dari data tersebut mempelajari hubungan antara variabel independen dan dependen, sementara kelompok pengujian difungsikan guna mengukur keandalan arsitektur yang telah dirancang. Skema Regresi Linier Berganda selanjutnya dilatih dengan memanfaatkan library Scikit-learn, model ini menghasilkan hubungan berbasis angka yang menghubungkan variabel bebas dan variabel terikat melalui suatu formula regresi, hal ini memungkinkan kita untuk melakukan estimasi nilai kendaraan secara lebih akurat.

Setelah tahap pelatihan, dilakukan evaluasi model menggunakan beberapa metrik, yaitu MSE, RMSE, MAE, serta  $R^2$ , hasil pengujian mengonfirmasi keandalan performa model dalam melakukan proyeksi nilai jual kendaraan bekas secara tepat. Selanjutnya, model yang telah dibangun diimplementasikan dalam bentuk antarmuka pengguna menggunakan Gradio, implementasi ini bertujuan untuk memudahkan pengguna dalam memperkirakan harga mobil bekas secara interaktif. Antarmuka sistem menyediakan beberapa input yang harus diisi oleh pengguna, seperti model, tahun, kilometer, bahan bakar, transmisi, pajak, cc, dan mpg. Setelah pengguna memasukkan data tersebut, sistem akan memproses input menggunakan model Regresi Linier Berganda dan menghasilkan output berupa nilai prediksi harga mobil bekas.

Hasil implementasi menunjukkan bahwa sistem berfungsi dengan baik dan mampu menghasilkan prediksi harga mobil bekas berdasarkan data input pengguna. Dengan adanya sistem ini, pengguna dapat memperoleh estimasi harga mobil bekas secara cepat, mudah, dan akurat. Implementasi antarmuka Gradio juga meningkatkan kemudahan penggunaan sistem karena pengguna tidak perlu memahami proses perhitungan yang digunakan dalam model prediksi. Dengan demikian, implementasi sistem prediksi harga mobil bekas berbasis metode Regresi Linier Berganda dengan antarmuka Gradio telah berhasil dikembangkan dan dapat dimanfaatkan sebagai alat bantu dalam memperkirakan harga mobil bekas berdasarkan spesifikasi kendaraan.

### 2.1.8 Evaluasi Model

Pengujian ini difokuskan pada pengukuran presisi serta efektivitas model Regresi Linier Berganda dalam menghasilkan estimasi harga pasar mobil bekas. Pengujian tersebut menggunakan beberapa metrik evaluasi, yaitu *MSE*, *RMSE*, *MAE*, serta  $R^2$ . Peneliti menerapkan beragam metrik evaluasi ini supaya kualitas prediksi yang dihasilkan oleh model dapat dibedah secara mendalam dari berbagai sisi.

Berikut merupakan metrik evaluasi yang akan digunakan:

#### a. Koefisien Determinasi ( $R^2$ )

Secara statistik,  $R^2$  menunjukkan seberapa besar variasi pada variabel terikat yang mampu dijelaskan oleh model regresi yang dibangun. Angka ini memiliki rentang nilai dari 0 hingga 1. Semakin dekat angka tersebut ke angka 1, berarti variabel bebas memiliki daya pengaruh yang lebih kuat terhadap variabel terikatnya, sehingga model regresi dianggap semakin baik [15].

$$R^2 = \frac{b_1(\sum x_1y) + b_2(\sum x_2y) + \dots + b_n(\sum x_ny)}{\sum y^2} \quad (2)$$

Pada persamaan tersebut,  $R^2$  bertindak sebagai koefisien determinasi yang menunjukkan persentase variabilitas harga mobil yang mampu dijelaskan oleh model. Notasi  $b_1$  hingga  $b_n$  merepresentasikan koefisien regresi parsial untuk setiap karakteristik teknis kendaraan. Sementara itu, komponen  $\sum x_ny$  menggambarkan jumlah produk deviasi antara variabel independen spesifik dan variabel dependen. Seluruh hasil akumulasi pada pembilang tersebut kemudian dibagi dengan  $\sum y^2$ , yang menunjukkan jumlah kuadrat total deviasi dari harga aktual kendaraan. Melalui rasio ini, tingkat akurasi dan kelayakan model regresi dapat dikuantifikasi secara objektif.

#### b. Mean Absolute Error (MAE)

*MAE* memberikan angka rata-rata kesalahan melalui perhitungan jarak absolut antara angka sebenarnya dengan angka prediksi, sehingga tingkat eror model bisa terlihat jelas [16].

$$MAE = \frac{1}{n} \sum_{t=1}^n |f_t - y_t| \quad (3)$$

Evaluasi tingkat kesalahan rata-rata dari model prediksi diukur menggunakan metrik Mean Absolute Error (MAE). Dalam perhitungannya, nilai MAE diperoleh dengan membagi hasil penjumlahan seluruh selisih absolut antara harga aktual dan harga prediksi dengan total sampel data mobil bekas ( $n$ ) di dalam dataset. Simbol nilai mutlak digunakan



untuk memastikan seluruh deviasi bernilai positif tanpa memedulikan arah kesalahannya. Selisih tersebut dihitung berdasarkan selisih antara variabel  $y_t$  yang merepresentasikan harga riil mobil Toyota di pasar dengan variabel  $f_t$  sebagai nilai harga prediksi yang dihasilkan oleh model regresi. Melalui akumulasi perbedaan absolut ini, indeks MAE secara objektif dapat menggambarkan seberapa jauh jarak rata-rata antara hasil estimasi model dengan kondisi harga pasar yang sebenarnya.

c. *Mean Squared Error (MSE)*

*Mean Squared Error (MSE)* memberikan gambaran mengenai tingkat kesalahan dengan cara merata-ratakan hasil kuadrat selisih antara output model dengan angka riil di lapangan [17].

$$MSE = \frac{1}{n} \sum_{n=1}^n (y_n - \hat{y}_n)^2 \quad (4)$$

Selain MAE, kinerja model dalam memproyeksikan harga kendaraan juga dievaluasi menggunakan metrik Mean Squared Error (MSE). Nilai MSE diperoleh dengan menghitung rata-rata dari seluruh akumulasi kuadrat selisih antara nilai aktual dan nilai prediksi. Di dalam kalkulasi tersebut, variabel  $y_n$  melambangkan harga riil atau aktual dari mobil bekas Toyota, sedangkan variabel  $\hat{y}_n$  merupakan nilai harga hasil estimasi yang dikeluarkan oleh model regresi. Selisih antara kedua nilai ini kemudian dikuadratkan sebelum dibagi dengan total sampel data ( $n$ ) yang diobservasi. Proses penguadratan ini secara statistik berfungsi memberikan bobot penalti yang lebih besar terhadap error prediksi berskala besar, sehingga hasil akhir MSE dapat menyajikan nilai varians kesalahan model dalam mengestimasi harga pasar secara akurat.

d. *Root Mean Squared Error (RMSE)*

*RMSE* berperan dalam mengukur jarak kesalahan antara data keluaran dan data asli dengan skema akar kuadrat. Kualitas prediksi dianggap semakin baik dan akurat saat nilai RMSE berada pada angka yang sekecil mungkin [18].

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

Sebagai metrik evaluasi terakhir, keakuratan model dalam mengestimasi harga jual kendaraan dinilai menggunakan Root Mean Squared Error (RMSE). Secara esensial, metrik ini menghitung akar kuadrat dari rata-rata akumulasi selisih kuadrat antara nilai aktual dan nilai prediksi. Pada prosesnya, nilai harga riil mobil Toyota di pasar disimbolkan oleh variabel  $y_i$ , sedangkan hasil proyeksi dari model regresi dinyatakan oleh variabel  $\hat{y}_i$ . Selisih dari kedua variabel tersebut dikuadratkan secara kumulatif untuk seluruh sampel data ( $n$ ), sebelum akhirnya dilakukan penarikan akar kuadrat pada hasil akhir. Penggunaan operasi akar kuadrat ini secara statistik berfungsi untuk mengembalikan satuan matriks kesalahan agar setara kembali dengan satuan mata uang asli dari harga mobil, sehingga menghasilkan nilai deviasi standar dari sisaan prediksi model yang lebih representatif.

### 2.1.9 Interpretasi

Tahap interpretasi dilakukan setelah model regresi linier berganda dibangun dan dievaluasi menggunakan metrik MSE. Tujuan tahap ini adalah memahami makna dari hasil analisis yang diperoleh, baik dari sisi statistik maupun konteks bisnis atau bidang penelitian.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Berdasarkan tahap pengumpulan data yang telah dirancang, diperoleh total data valid sebanyak 400 baris data dari repositori Kaggle dan forum jual beli lokal Pekanbaru. Data yang berhasil dikumpulkan ini mencakup rentang tahun kendaraan dari 2005 hingga 2025 dengan variasi model kendaraan merek Toyota yang sangat dinamis. Seluruh 400 baris data ini telah memenuhi kelengkapan delapan atribut independen dan satu atribut dependen, sehingga dinyatakan siap untuk masuk ke tahap pra-proses dan pemodelan tanpa adanya data yang kosong.

### 3.2 Preprocessing Data

Seluruh tahapan pra-pemrosesan data terdiri dari 400 data mobil bekas dikerjakan melalui lingkungan kerja Google Colab dengan rincian hasil sebagai berikut:

a. *Pembersihan Data (Data Cleaning)* :

Berdasarkan hasil pengecekan pada Google Colab, tidak ditemukan adanya data duplikat pada 400 dataset yang digunakan. Namun, ditemukan beberapa data dengan nilai kosong (missing value), sehingga dilakukan penghapusan baris data tersebut untuk menjaga validitas model.

b. *Transformasi Data (Data Transformation)*

Variabel kategorikal berupa tipe transmisi dan jenis bahan bakar telah berhasil ditransformasikan menjadi bentuk numerik biner (0 dan 1) melalui teknik One Hot Encoding. Seluruh data kategorikal kini telah siap digunakan sebagai input matriks pada rumus Regresi Linier Berganda.

c. *Normalisasi Data (Data Normalization)*



Normalisasi dilakukan menggunakan metode MinMaxScaler pada variabel numerik seperti tahun, kilometer, pajak, mpg, dan cc. Hasil dari normalisasi data ditampilkan pada Tabel 2.

**Tabel 2.** Hasil Normalisasi Data

No	Tahun	Kilometer	Pajak	Mpg	Cc
0	0.450	0.254	0.159	0.433	0.000
1	0.500	0.193	0.165	0.436	0.000
2	0.500	0.309	0.165	0.436	0.000
3	0.500	0.302	0.165	0.436	0.000
4	0.500	0.222	0.165	0.436	0.000

Berdasarkan rekaman data yang tersaji pada Tabel 2, dapat dirinci bahwa seluruh atribut numerik kendaraan kini telah berhasil dipetakan ke dalam rentang nilai minimal 0 dan maksimal 1 setelah melalui transformasi MinMaxScaler. Sebagai ilustrasi pada sampel baris ke-0, nilai tahun perakitan kendaraan yang semula berupa angka tahun ribuan bertransformasi menjadi 0,450, atribut kilometer berada di angka 0,254, dan aspek perpajakan tercatat sebesar 0,159. Sementara itu, representasi angka 0,000 pada kolom kapasitas mesin (cc) di seluruh sampel Tabel 2 bukan menunjukkan ketiadaan nilai atau data kosong. Angka 0,000 tersebut merupakan nilai batas bawah (minimum) yang mengonfirmasi bahwa unit kendaraan dalam sampel ini kebetulan memiliki kapasitas silinder mesin paling kecil di dalam keseluruhan dataset (misalnya varian mobil 1.000 cc). Melalui standardisasi linear yang terperinci pada Tabel 2 ini, ketimpangan satuan ukuran antar-fitur berhasil dieliminasi secara total sebelum memasuki tahapan fitting model regresi.

#### d. Pembagian Data (Data Splitting)

Dataset mobil bekas merek Toyota dipisahkan menjadi kelompok data latih dan data uji melalui lingkungan kerja Google Colab. Rincian hasil pembagian serta proporsi dataset dari beberapa skenario pengujian ditampilkan pada Tabel 3.

**Tabel 3.** Pembagian Dataset

No	Pembagian Data	
	Data Training	Data Testing
1	280 data (70%)	120 data (30%)
2	320 data (80%)	80 data (20%)
3	360 data (90%)	40 data (10%)

Tabel 3 menyajikan beberapa skenario pembagian dataset yang digunakan dalam penelitian, yaitu rasio 70:30, 80:20, dan 90:10 antara data latih dan data uji. Setiap skenario menunjukkan proporsi data yang digunakan untuk proses pelatihan dan pengujian model dari total 400 data yang tersedia. Dalam penelitian ini, dipilih skenario 80:20, dengan 320 data sebagai data latih dan 80 data sebagai data uji, karena dinilai memberikan keseimbangan yang optimal antara proses pelatihan dan evaluasi model. Jika ditinjau lebih rinci berdasarkan data pada Tabel 3 tersebut, pemilihan rasio ini didasarkan pada pertimbangan karakteristik komputasi dari setiap skenario. Pada skenario rasio 70:30, alokasi data latih yang lebih sedikit (280 data) memang dapat memicu peningkatan akurasi training yang sangat tinggi pada proses seleksi variabel berikutnya karena beban data yang dipelajari model lebih kecil. Namun, porsi data uji yang terlalu besar berisiko mengurangi stabilitas generalisasi model. Sebaliknya, pada skenario rasio 90:10, jumlah data uji menyusut menjadi hanya 40 data, sehingga dinilai kurang objektif dalam menggambarkan performa asli model akibat keterbatasan sampel pengujian. Oleh karena itu, pengujian pada skenario 80:20 dipilih sebagai model final karena mampu menjamin ketersediaan data latih yang cukup bagi algoritma untuk mengenali pola harga pasar secara mendalam, sekaligus mempertahankan volume data uji yang valid untuk mengukur akurasi prediksi secara stabil.

### 3.3 Seleksi Variabel

Proses seleksi variabel dilakukan dengan membandingkan nilai ( $R^2$ ) yang dihasilkan pada data training dan data testing dari beberapa kombinasi variabel independen. Melalui pendekatan ini diharapkan diperoleh kombinasi variabel yang menghasilkan performa prediksi yang optimal untuk model Regresi Linier Berganda [13]. Untuk memperoleh hasil evaluasi yang lebih terukur, Kombinasi variabel ini diterapkan pada tiga skenario pembagian dataset, yaitu 70:30, 80:20, dan 90:10 guna menilai konsistensi kinerja model terhadap variasi pembagian data latih dan data uji. Kinerja dari setiap kombinasi variabel dinilai berdasarkan capaian nilai  $R^2$  yang paling tinggi mendekati angka 1 serta memiliki selisih paling kecil antara data latih dan data uji untuk memastikan model terhindar dari fenomena overfitting [13]. Pemilihan variabel final dalam penelitian ini disesuaikan berdasarkan tingkat relevansi tertinggi masing-masing atribut terhadap fluktuasi harga mobil bekas Toyota. Adapun rincian hasil pengujian performa dari beberapa skenario kombinasi variabel tersebut disajikan secara terstruktur pada Tabel 4.

**Tabel 4.** Seleksi Variabel dengan Pembagian Dataset Mobil 70: 30

model	tahun	transmisi	kilometer	bahan bakar	pajak	mpg	cc	$R^2$ Train	$R^2$ Test
✓	✓	✓	✓	✓	✓	✓	✓	99%	96%



model	tahun	transmisi	kilometer	bahan bakar	pajak	mpg	cc	R <sup>2</sup> Train	R <sup>2</sup> Test
✓	✓	X	X	X	X	✓	✓	98%	95%
✓	✓	X	✓	X	X	✓	✓	98%	95%
✓	✓	X	X	X	X	X	✓	98%	93%
✓	✓	X	X	X	X	✓	X	97%	92%
✓	✓	X	X	✓	X	X	X	98%	92%

Berdasarkan analisis hasil pengujian pada Tabel 4 dengan skenario rasio pembagian data 70:30, sebagian besar kombinasi variabel independen terbukti menghasilkan performa model yang kuat. Pola ini ditunjukkan oleh perolehan nilai R<sup>2</sup> training yang konsisten berada di rentang tinggi antara 98% hingga 99%, serta nilai R<sup>2</sup> testing yang bergerak stabil pada rentang 92% hingga 96%. Jika ditinjau secara terperinci pada Tabel 4, penggunaan kombinasi fitur yang lebih ringkas seperti keterlibatan komponen tahun, cc, dan mpg sudah mampu menunjukkan kemampuan generalisasi yang stabil dengan risiko overfitting yang rendah karena selisih antara performa training dan testing relatif kecil. Namun, ketika model dikonfigurasi menggunakan kombinasi fitur yang paling lengkap (meliputi model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc), model regresi menghasilkan nilai R<sup>2</sup> training tertinggi mencapai 99% dengan tingkat akurasi data testing sebesar 96%. Hasil evaluasi pada Tabel 4 ini secara empiris mengonfirmasi bahwa algoritma regresi linier berganda telah berhasil memetakan keterkaitan linear dan interaksi kompleks di antara seluruh faktor spesifikasi teknis kendaraan terhadap target log-harga secara efektif dan akurat.

**Tabel 5.** Seleksi Variabel dengan Pembagian Dataset Mobil 80: 20

model	tahun	transmisi	kilometer	bahan bakar	pajak	mpg	cc	R <sup>2</sup> Train	R <sup>2</sup> Test
✓	✓	✓	✓	✓	✓	✓	✓	98%	96%
✓	✓	X	X	X	X	✓	✓	98%	95%
✓	✓	X	✓	X	X	✓	✓	98%	95%
✓	✓	X	X	X	X	X	✓	98%	93%
✓	✓	X	X	X	X	✓	X	97%	91%
✓	✓	X	X	✓	X	X	X	98%	91%

Selanjutnya, hasil analisis eksperimen pada Tabel 5 dengan skenario rasio pembagian data 80:20 menunjukkan bahwa sebagian besar kombinasi variabel independen menghasilkan performa pemodelan yang sangat kuat dan konsisten. Karakteristik ini dibuktikan oleh perolehan nilai R<sup>2</sup> training yang stabil di angka 97% hingga 98%, serta capaian nilai R<sup>2</sup> testing yang bergerak pada rentang 91% hingga 96%. Jika ditinjau secara terperinci pada Tabel 5, konfigurasi model yang menggunakan kombinasi variabel paling lengkap meliputi model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc berhasil mencatatkan nilai R<sup>2</sup> testing tertinggi sebesar 96% dengan nilai R<sup>2</sup> training sebesar 98%. Di sisi lain, pengujian menggunakan kombinasi variabel yang lebih sederhana, seperti hanya melibatkan fitur model, tahun, dan mpg, menghasilkan nilai R<sup>2</sup> training yang sedikit lebih rendah yaitu 97% dengan capaian R<sup>2</sup> testing sebesar 91%. Secara keseluruhan, visualisasi data pada Tabel 5 ini membuktikan bahwa arsitektur regresi linier berganda tetap mampu menangkap hubungan linear yang kuat dan kompleks antara seluruh variabel independen dengan target log-harga secara optimal.

**Tabel 6.** Seleksi Variabel dengan Pembagian Dataset Mobil 90: 10

model	tahun	transmisi	kilometer	bahan bakar	pajak	mpg	cc	R <sup>2</sup> Train	R <sup>2</sup> Test
✓	✓	X	X	X	X	✓	✓	98%	94%
✓	✓	X	✓	X	X	✓	✓	98%	94%
✓	✓	✓	✓	✓	✓	✓	✓	98%	94%
✓	✓	X	X	X	X	X	✓	97%	91%
✓	✓	X	X	X	X	✓	X	97%	86%
✓	✓	X	X	✓	X	X	X	97%	85%

Sementara itu, hasil pengujian pada Tabel 6 dengan skenario rasio pembagian dataset 90:10 menunjukkan bahwa sebagian besar kombinasi variabel independen menghasilkan nilai R<sup>2</sup> training yang tinggi, yaitu berada di rentang antara 97% hingga 98%, dengan capaian nilai R<sup>2</sup> testing yang bergerak antara 85% hingga 94%. Jika ditinjau secara terperinci pada Tabel 6, penggunaan kombinasi variabel yang paling lengkap meliputi komponen model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc mampu memberikan nilai R<sup>2</sup> testing tertinggi sebesar 94% dengan nilai R<sup>2</sup> training sebesar 98%. Di sisi lain, pengujian dengan konfigurasi variabel yang lebih terbatas, seperti hanya menggunakan kombinasi fitur model, tahun, dan bahan bakar, menghasilkan nilai R<sup>2</sup> testing sebesar 85%. Meskipun terjadi variasi penurunan angka akurasi pada data uji, secara umum model regresi linier berganda pada skenario ini tetap mampu menangkap hubungan linear antara fitur-fitur kendaraan dengan target log-harga secara baik. Melalui seluruh rangkaian eksperimen yang didokumentasikan pada Tabel 6 ini, dapat disimpulkan bahwa kombinasi variabel utama yang meliputi model, tahun,



transmisi, kilometer, bahan bakar, mpg, dan cc secara konsisten memegang peranan paling berpengaruh dalam menghasilkan akurasi prediksi harga mobil yang optimal.

Berdasarkan hasil pada Tabel 4, Tabel 5, dan Tabel 6, kombinasi semua variabel independen, yaitu model, tahun, transmisi, kilometer, bahan bakar, pajak, MPG, dan CC, secara konsisten memberikan nilai  $R^2$  yang tinggi pada data training maupun testing. Nilai  $R^2$  yang relatif tinggi dan selisih kecil antara training dan testing menunjukkan bahwa model mampu menjelaskan variasi data secara menyeluruh sekaligus memiliki kemampuan prediksi yang baik. Dengan menggunakan semua variabel, model regresi linier berganda dapat menangkap hubungan kompleks antara fitur-fitur kendaraan dan harga mobil, sehingga menghasilkan prediksi yang akurat dan stabil. Hal ini menunjukkan bahwa penggunaan seluruh variabel independen penting untuk meningkatkan kinerja model, dibandingkan hanya menggunakan sebagian variabel saja.

### 3.4 Implementasi Regresi Linier Berganda

Setelah melalui tahap pra-pemrosesan data, model Regresi Linier Berganda berhasil dibangun menggunakan transformasi logaritma pada variabel harga serta normalisasi data menggunakan MinMaxScaler. Berdasarkan hasil pelatihan model pada lingkungan kerja Google Colab, diperoleh nilai intercept ( $\beta_0$ ) sebesar 17,7669 beserta nilai koefisien untuk masing-masing variabel prediktor. Berikut merupakan persamaan model regresi linear berganda yang diperoleh:

$$\log(\text{harga}) = 17,7669 + 1,4452X_1 - 0,1408X_2 + 0,0501X_3 - 0,3325X_4 + 0,3556X_5$$

Melalui rumus tersebut, tahapan sistemis algoritma dalam menyelesaikan masalah penaksiran nilai aset dapat dibaca secara transparan berdasarkan fungsi masing-masing simbolnya. Nilai intercept ( $\beta_0 = 17,7669$ ) bertindak sebagai titik acuan konstan atau nilai dasar awal dari nilai  $\log(\text{harga})$  ketika seluruh variabel prediktor berada pada kondisi nol setelah melewati proses normalisasi data. Selanjutnya, arah pengaruh dari karakteristik kendaraan ditentukan oleh orientasi nilai koefisien di depan simbol prediktornya. Simbol prediktor dengan koefisien positif, yang meliputi variabel tahun ( $X_1$ ) beban pajak kendaraan ( $X_3$ ), dan cc ( $X_5$ ), menunjukkan adanya hubungan yang berbanding lurus dengan nilai jual, di mana setiap eskalasi nilai pada fitur-fitur teknis tersebut secara otomatis akan mendongkrak estimasi harga mobil. Sebaliknya, simbol dengan koefisien negatif pada variabel kilometer ( $X_2$ ) dan mpg ( $X_4$ ) mengonfirmasi hubungan yang berbanding terbalik, yang berarti semakin tinggi akumulasi angka kilometer atau konsumsi bahan bakarnya, maka proyeksi harga yang dikeluarkan oleh model regresi akan mengalami penurunan secara signifikan.

Pada data mobil Toyota Innova tahun 2012 dengan kilometer 103.000, pajak Rp3.100.000, cc 2.5, transmisi matic, dan bahan bakar diesel, nilai prediksi diperoleh dengan terlebih dahulu melakukan normalisasi data, kemudian dimasukkan ke dalam persamaan regresi untuk menghasilkan nilai  $\log(\text{harga})$ . Selanjutnya, nilai tersebut dikembalikan ke bentuk harga sebenarnya sehingga diperoleh estimasi harga dalam satuan rupiah.

Model regresi linear berganda tersebut selanjutnya diimplementasikan pada data testing, dan sampel hasilnya ditunjukkan pada Tabel 7 berikut:

**Tabel 7.** Hasil Prediksi

No	Aktual	Prediksi
0	19.115	19.081
1	19.096	19.057
2	20.367	20.441
3	19.354	19.391
4	19.297	19.316

Berdasarkan hasil pada Tabel 7, nilai prediksi yang dihasilkan model menunjukkan kedekatan yang tinggi dengan nilai aktual dalam bentuk logaritma harga. Selisih antara nilai aktual dan prediksi relatif kecil, yang mengindikasikan bahwa model memiliki tingkat akurasi yang baik dalam memodelkan hubungan antar variabel. Capaian ini diperkuat oleh nilai koefisien determinasi ( $R^2$ ) yang sudah mendekati angka satu, model telah mampu menjelaskan sebagian besar variasi data dengan cukup baik. Hal ini menunjukkan bahwa penggunaan seluruh variabel independen dalam model regresi linear berganda mampu meningkatkan kinerja prediksi secara keseluruhan.

### 3.5 Implementasi Antarmuka

Setelah model berhasil dibangun, langkah selanjutnya adalah mengimplementasikan model dalam bentuk antarmuka pengguna menggunakan Gradio. Implementasi ini bertujuan untuk memudahkan pengguna dalam melakukan prediksi harga mobil bekas secara interaktif. Antarmuka yang dibangun menyediakan beberapa input yang harus diisi oleh pengguna, yaitu model, tahun, kilometer, bahan bakar, transmisi, pajak, mpg, dan cc. Setelah pengguna memasukkan data tersebut, sistem akan memproses input menggunakan model Regresi Linier Berganda yang telah dilatih sebelumnya.

Sistem kemudian menghasilkan output berupa nilai prediksi harga mobil bekas berdasarkan data input yang diberikan oleh pengguna. Implementasi antarmuka menggunakan Gradio memungkinkan sistem dapat digunakan secara mudah dan interaktif tanpa memerlukan pengetahuan teknis mengenai pemrograman atau machine learning. Hasil implementasi menunjukkan bahwa sistem dapat berjalan dengan baik dan mampu menghasilkan prediksi harga mobil bekas secara cepat dan akurat. Dengan adanya antarmuka ini, pengguna dapat memperoleh estimasi harga mobil bekas berdasarkan spesifikasi kendaraan yang dimasukkan.



Pada sistem prediksi harga mobil bekas ini, arsitektur antarmuka halaman utama dirancang untuk memisahkan hak akses antara pengguna umum dan pengelola data. Representasi visual dari tata letak halaman utama beserta komponen-komponennya disajikan secara jelas pada Gambar 3.

Berdasarkan struktur visual yang ditunjukkan pada Gambar 3 tersebut, halaman Home dan menu Prediksi dapat diakses langsung oleh konsumen secara terbuka tanpa harus melalui proses login. Halaman Prediksi tersebut digunakan oleh konsumen untuk memperkirakan estimasi harga jual kendaraan berdasarkan data spesifikasi yang dimasukkan. Sementara itu, komponen formulir Login Admin yang terletak pada bilah samping (sidebar) Gambar 3, dirancang secara khusus untuk administrator. Fitur tersebut mewajibkan pengisian username dan password yang valid agar administrator dapat masuk ke dalam sistem pengolahan database untuk mengelola, menambah, atau menghapus data mobil bekas. Melalui pengorganisasian elemen pada Gambar 3, alur interaksi pengguna dengan sistem dapat berjalan dengan lebih terstruktur dan aman.

Sistem Prediksi Harga Mobil Bekas (Regresi Linier Berganda)

Home Selamat Datang di Sistem Showroom

Data

Tambah Mobil Solusi Cerdas Prediksi Harga Mobil Bekas

Prediksi

Sistem ini dirancang untuk membantu Anda mengetahui estimasi harga jual atau beli mobil secara objektif.

Keunggulan Sistem Kami:

- Cepat & Praktis Masukkan spesifikasi mobil (CC, Tahun, Pajak, dll) dan dapatkan estimasi harga dalam hitungan detik.
- Transparansi Data (MSE & Error) Kami menyediakan nilai MSE (Mean Squared Error) sehingga Anda tahu tingkat keakuratan setiap prediksi.
- Berdasarkan Data Real Prediksi dihitung menggunakan algoritma "Linear Regression" yang memprediksi harga mobil bekas.

Login Admin

Username  
admin

Password  
\*\*\*\*\*

Login

Status  
Login Admin Berhasil

Cara Mulai:  
Klik tombol **Prediksi** di menu samping, masukkan data mobil Anda, dan lihat hasilnya langsung!

Tips: Pastikan data Kilometer dan Tahun akurat agar estimasi error tetap rendah.

Silakan pilih menu di samping kiri untuk memulai. Jika Anda adalah pengelola, silakan Login untuk menambah atau menghapus data koleksi.

**Gambar 3.** Antarmuka Halaman Home Sistem Prediksi Harga Mobil Bekas

Selanjutnya, Halaman prediksi digunakan untuk memperkirakan harga mobil bekas berdasarkan atribut kendaraan yang dimasukkan pengguna. Data tersebut diproses menggunakan model Regresi Linier Berganda untuk menghasilkan estimasi harga mobil, sebagaimana ditunjukkan pada Gambar 4.

Sistem Prediksi Harga Mobil Bekas (Regresi Linier Berganda)

Home Prediksi Harga Mobil

Data

Tambah Mobil

Prediksi

Login Admin

Username  
admin

Password  
\*\*\*\*\*

Login

Status  
Login Admin Berhasil

Tahun

Kilometer

Pajak

MPG

CC

Model

Transmisi

**Gambar 4.** Antarmuka Halaman Prediksi Harga Mobil Bekas

Berdasarkan tampilan antarmuka yang ditunjukkan pada Gambar 4 tersebut, halaman prediksi digunakan secara khusus untuk memperkirakan estimasi harga jual kendaraan berdasarkan beberapa atribut spesifikasi teknis yang dimasukkan oleh pengguna. Seluruh data atribut masukan tersebut kemudian diproses secara langsung di dalam sistem menggunakan model Regresi Linier Berganda. Hasil komputasi dari algoritma ini akan memicu sistem pada Gambar 4



untuk menghasilkan taksiran nominal harga mobil secara instan, akurat, dan objektif yang langsung ditampilkan pada panel hasil prediksi.

Setelah administrator berhasil masuk melalui menu login pada Gambar 3, sistem akan mengarahkan admin ke Halaman Data Mobil dan Halaman Tambah Mobil. Pada kedua halaman ini, admin dapat memperbarui dataset kendaraan, baik dengan menambahkan data mobil baru beserta spesifikasinya maupun menghapus data yang salah input. Seluruh perubahan tersebut akan otomatis tersimpan ke dalam file Excel yang digunakan sebagai sumber data (dataset) utama untuk proses prediksi harga.

### 3.6 Evaluasi Model

Proses evaluasi model dilakukan untuk mengukur sejauh mana performa dan tingkat akurasi algoritma Regresi Linier Berganda dalam mengestimasi nilai jual kendaraan. Berdasarkan hasil pengujian yang dilakukan pada berbagai skenario pemodelan, performa akhir dari model regresi dinilai berdasarkan empat metrik utama, yaitu Koefisien determinan ( $R^2$ ), Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE). Rincian capaian nilai metrik evaluasi tersebut disajikan pada Tabel 8.

**Tabel 8.** Evaluasi Model

Model	$R^2$	MAE	MSE	RMSE
Training	0.985	0.046	0.004	0.062
Testing	0.955	0.071	0.010	0.100

Nilai evaluasi model diperoleh dengan memanfaatkan seluruh variabel independen dalam membangun model regresi linier berganda, dengan tujuan untuk menilai kinerja model secara komprehensif tanpa melalui proses seleksi variabel. Pendekatan ini memungkinkan setiap variabel memberikan kontribusi terhadap proses prediksi harga mobil bekas, sehingga hasil evaluasi mencerminkan hubungan antara variabel independen dan variabel dependen secara menyeluruh, berdasarkan hasil pengujian, diperoleh nilai MAE sebesar 0,046 pada data training dan 0,071 pada data testing. Hal ini menunjukkan bahwa rata-rata selisih absolut antara nilai prediksi dan nilai aktual relatif kecil. Selanjutnya, nilai MSE sebesar 0,004 pada data training dan 0,010 pada data testing mengindikasikan bahwa rata-rata kesalahan model juga tergolong rendah, sementara itu, RMSE sebesar 0,062 pada data training dan 0,100 pada data testing memberikan gambaran tingkat kesalahan prediksi dalam skala yang lebih mudah diinterpretasikan. Perbedaan nilai evaluasi antara data training dan testing yang relatif kecil menunjukkan bahwa model memiliki kemampuan generalisasi yang baik serta tidak mengalami overfitting secara signifikan, selain itu, nilai  $R^2$  sebesar 0,985 pada data training dan 0,955 pada data testing menunjukkan bahwa model mampu menjelaskan sekitar 95,5% variasi harga mobil bekas berdasarkan variabel independen yang digunakan.

### 3.7 Interpretasi

Hasil pengujian model regresi linier berganda menunjukkan performa yang sangat tinggi dengan perolehan nilai  $R^2$  sebesar 0,985 pada data training dan 0,955 pada data testing. Nilai tersebut mengindikasikan bahwa fitur model, tahun, transmisi, kilometer, bahan bakar, pajak, mpg, dan cc mampu menjelaskan 95,5% variasi harga mobil bekas secara akurat. Rendahnya nilai metrik error, seperti MAE sebesar 0,071, MSE sebesar 0,010 dan RMSE sebesar 0,100, memperkuat tingkat ketepatan model dalam melakukan estimasi harga di lapangan. tingkat presisi model dipertegas dengan rendahnya nilai metrik kesalahan. Nilai MAE (Mean Absolute Error) sebesar 0,071 menunjukkan bahwa rata-rata selisih absolut antara harga prediksi dan harga aktual sangat tipis. Penggunaan MSE (Mean Squared Error) sebesar 0,010 dan RMSE (Root Mean Squared Error) sebesar 0,100 memberikan indikasi bahwa model tidak memiliki kesalahan prediksi yang tinggi secara signifikan. Nilai RMSE yang rendah memiliki signifikan yang tinggi dalam konteks harga mobil bekas, karena menunjukkan bahwa estimasi yang dihasilkan model cukup aman dan stabil untuk digunakan sebagai acuan transaksi di lapangan tanpa risiko selisih harga yang merugikan salah satu pihak. selisih performa yang minimal antara data training dan testing juga membuktikan bahwa model memiliki kestabilan prediksi yang baik serta terhindar dari masalah overfitting. Dibandingkan dengan penelitian terdahulu oleh [7] yang memperoleh  $R^2$  sebesar 0,94, model yang dikembangkan dalam penelitian ini menawarkan peningkatan akurasi sebesar 1,5%. Peningkatan ini, meski terlihat kecil secara angka, memberikan dampak signifikan pada nilai yang diprediksi, mengingat harga kendaraan memiliki rentang nominal yang besar. Keunggulan ini kemungkinan besar dipengaruhi oleh kualitas pra-pemrosesan data (data preprocessing) dan pemilihan fitur yang lebih relevan dalam menangkap karakteristik harga pasar saat ini. Secara fungsional, hasil ini memvalidasi bahwa penggunaan algoritma regresi linier berganda masih sangat relevan dan kuat untuk kasus prediksi harga yang memiliki hubungan linier antara fitur dan target. Model ini dapat bertindak sebagai instrumen pendukung keputusan yang objektif, membantu calon pembeli maupun penjual dalam menentukan harga pasar yang adil berdasarkan spesifikasi teknis kendaraan, sehingga meminimalkan perkiraan dalam proses negosiasi.

## 4. KESIMPULAN

Berdasarkan hasil penelitian, metode Regresi Linier Berganda terbukti efektif dalam memprediksi harga mobil bekas dengan tingkat akurasi yang baik. Model dikembangkan menggunakan beberapa variabel independen, yaitu tahun,



kilometer, pajak, mpg, dan cc. Proses penelitian mencakup tahap preprocessing data, yang meliputi pembersihan data, transformasi variabel kategorikal menggunakan One Hot Encoding, serta normalisasi data melalui MinMaxScaler. Dataset kemudian dibagi menjadi satu skenario evaluasi dengan tujuan menilai performa model secara menyeluruh. Model dievaluasi menggunakan empat metrik utama, yaitu  $R^2$ , MAE, MSE, dan RMSE. Hasil evaluasi menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang rendah, ditunjukkan oleh nilai MSE pada data training sebesar 0,004 dan data testing sebesar 0,010, MAE pada data training sebesar 0,046 dan data testing sebesar 0,071, dan RMSE pada data training sebesar 0,062 dan sebesar 0,100 pada data testing. Selain itu,  $R^2$  pada data training sebesar 0,985 dan pada data testing sebesar 0,955 menandakan bahwa model mampu menjelaskan sekitar 95,5% variasi harga mobil bekas berdasarkan variabel yang digunakan. Model yang dikembangkan selanjutnya diimplementasikan dalam antarmuka pengguna menggunakan Gradio, sehingga memungkinkan prediksi harga mobil bekas secara interaktif dan mudah diakses. Penelitian ini memberikan kontribusi dalam pengembangan sistem prediksi harga kendaraan berbasis data, yang dapat membantu pengguna memperoleh estimasi harga secara objektif. Untuk penelitian berikutnya, model dapat ditingkatkan dengan menambahkan jumlah dataset yang lebih besar dan mempertimbangkan penerapan metode machine learning lain guna meningkatkan akurasi prediksi.

## REFERENCES

- [1] A. S. Bacht, A. T. Agustama, H. Wiranata, E. Daniati, and A. Ristyawan, "Prediksi Harga Mobil Toyota Bekas Menggunakan Algoritma Regresi Linear Berganda," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.*, vol. 8, no. 1, pp. 173–183, 2024, doi: 10.29407/inotek.v8i1.4924.
- [2] D. Miftahul Huda, G. Dwilestari, and A. R. Rinaldi, "Prediksi Harga Mobil Bekas Menggunakan Algoritma Regresi Linear Berganda," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 150–157, 2024, doi: 10.36499/jinrpl.v6i1.10266.
- [3] Ilham Amansyah, Jamaludin Indra, Euis Nurlaelasari, and Ayu Ratna Juwita, "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 4, pp. 1199–1216, 2024, doi: 10.31004/innovative.v4i4.12735.
- [4] P. K. Dewi and Yoviapridiansyah, "Penerapan Algoritma Linear Regression dalam Prediksi Harga Mobil Bekas Berbasis Machine Learning," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 5, no. 2, pp. 473–480, 2026, doi: 10.53513/jursi.v5i2.12494.
- [5] E. Dewi, S. Mulyani, F. Mulady, D. Ramadhan, A. Ariyantono, and D. Ramdani, "Estimasi Harga Jual Mobil Bekas Menggunakan Metode Regresi Linier Berganda," *e-Jurnal JUSITI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 9, no. 1, pp. 1–8, 2020, doi: 10.36774/jusiti.v9i1.649.
- [6] Y. Haliza, Putri, R. Rafiza, and J. Simanullang, "Penerapan Regresi Linier Berganda dalam Memprediksi IPM Berdasarkan Faktor Ekonomi dan Sosial di Sumatera Barat," *Socius J. Penelit. Ilmu-Ilmu Sos.*, vol. 2, no. 12, pp. 544–554, 2025, doi: 10.5281/zenodo.15696487.
- [7] P. Aulia Azhar, M. Arya Pratama, and R. Fitriani, "Prediksi Harga Mobil Audi Bekas Menggunakan Model Regresi Linear dengan Framework Streamlit," *J. Technol. Informatics*, vol. 6, no. 1, pp. 22–28, 2024, doi: 10.37802/joti.v6i1.763.
- [8] C. Whenjaya, H. Susanto, and Herman, "Model Prediksi Harga Mobil Bekas Menggunakan Regresi Linear Berganda Berdasarkan Fitur Kendaraan," *J. Digit. Ecosyst. Nat. Sustain.*, vol. 4, no. 2, pp. 95–99, 2024, doi: 10.63643/jodens.v4i2.263.
- [9] Nusyirwan, Irfadli, and Menhard, "Apakah Bauran Pemasaran Berpengaruh Terhadap Keputusan Pembelian Mobil Pada Agung Toyota Cabang Duri," *J. Ekon. Akunt. dan Manaj. Nusant.*, vol. 4, no. April, pp. 403–408, 2026, doi: 10.55338/jeama.v4i3.456.
- [10] A. Sajad and N. Nurmalitasari, "Analisis Faktor–Faktor Yang Mempengaruhi Harga Mobil Bekas Menggunakan Metode Regresi Linier," *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, pp. 266–270, 2023, [Online]. Available: <https://www.ojs.udb.ac.id/Senatib/article/view/3195>
- [11] N. O. Idris and F. Pontooyo, "Evaluasi Model Machine Learning untuk Prediksi Harga Mobil dengan Perbandingan Ensemble dan Regresi Linear," *J. Ilmu Komput. dan Sist. Inf.*, vol. 4, no. 1, pp. 129–143, 2025, doi: 10.70340/jirsi.v4i1.181.
- [12] M. Irpan, U. Rio, Karpen, and Hamdani, "Model Prediksi Jumlah Produksi Kelapa Sawit Menggunakan Regresi Linear Berganda Di PT . Surya Argolika Reksa," *J. Ilm. BETRIK*, vol. 16, no. 02, pp. 122–130, 2025, doi: 10.36050/dv9sd116.
- [13] I. J. Matondang, E. Budianita, F. Syafria, and I. Afrianty, "BULLETIN OF COMPUTER SCIENCE RESEARCH Sistem Prediksi Produksi Kelapa Sawit Berbasis Gradio Menggunakan Algoritma Regresi Linear Berganda," *Bull. Comput. Sci. Res.*, vol. 6, no. 2, pp. 664–672, 2026, doi: 10.47065/bulletincsr.v6i2.994.
- [14] M. Imam and N. Pasaribu, "Implementasi Data Mining Dengan Metode Regresi Linier Untuk Memprediksi Laju Pertumbuhan Penduduk Kota Binjai Implementation of Data Mining Using Linear Regression Method to Predict Population Growth Rate of Binjai City," *J. Ilmu Komput. dan Sist. Inf.*, no. September, pp. 405–415, 2025, [Online]. Available: <https://jurnal.unity-academy.sch.id/index.php/jirsi/article/view/247>
- [15] A. A. Asriadi, C. A. Anggraini, F. I. Puspita, H. Juan, M. A. Imbiri, and U. Papua, "Prediksi Partisipasi Jemaat dalam Ibadah Penutupan Tahun 2025 Menggunakan Metode Regresi Linier Sederhana: Studi Kasus di Jemaat Efrata," *Sci-Tech J.*, vol. 4, no. 2, pp. 246–256, 2025, doi: 10.56709/stj.v4i3.809.
- [16] M. U. Riyal *et al.*, "Analisa Perbandingan Metode Trend Moment dan Regresi Linear dalam Prediksi Kurs Mata Uang Rupiah terhadap," *J. Comput. Syst. Informatics*, vol. 6, no. 3, pp. 563–571, 2025, doi: 10.47065/josyc.v6i3.7400.
- [17] F. F. Az-zahra, T. Informatika, U. M. Sukabumi, K. Sukabumi, R. L. Berganda, and A. Regresi, "Prediksi hasil panen wortel menggunakan algoritma regresi linear berganda," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 10255–10262, 2024, doi: 10.36040/jati.v8i5.10954.
- [18] A. J. Karlina, M. Irsyad, F. Insani, and E. P. Cynthia, "Estimasi Hasil Panen Ayam Pedaging Menggunakan Algoritma Regresi Linear Berganda," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 966–976, 2023, doi: 10.30865/klik.v3i6.920.