

Prediksi Harga Mobil Toyota Bekas Menggunakan Algoritma Regresi Linear Berganda

Diterima: 10 Juni 2024
Revisi: 10 Juli 2024
Terbit: 1 Agustus 2024

^{1*}Achmad Syauqi Bacht, ²Andri Tri Agustama, ³Hadi Wiranata,
⁴Erna Daniati, ⁵Aidina Ristyawan
¹⁻⁵Universitas Nusantara PGRI Kediri
¹syauqi.bakti@gmail.com, ²andriagus060822@gmail.com,
³hadiwiranata311@gmail.com, ⁴ernadaniati@unpkediri.ac.id,
⁵adinaristi@unpkediri.ac.id

Abstrak—Kurangnya informasi tentang harga mobil bekas dapat menjadi hambatan bagi masyarakat dalam melakukan transaksi jual beli kendaraan karena mereka mungkin tidak memahami harga pasar yang sesuai. Oleh karena itu, penting untuk memiliki informasi yang akurat dan dapat diandalkan untuk mengetahui perkiraan harga yang dapat digunakan sebagai referensi. Dalam rangka mengatasi ini, penggunaan algoritma regresi linear berganda dapat membantu dalam memprediksi harga mobil bekas. Dalam penelitian ini, algoritma regresi linear berganda digunakan untuk memprediksi harga mobil bekas. Data yang digunakan untuk membangun model prediksi berasal dari repositori Kaggle. Proses pengujian model dilakukan dengan membagi data menjadi data pelatihan (90%) dan data pengujian (10%). Setelah proses pelatihan, model prediksi didapatkan dengan nilai rata-rata harga sebesar 1512.262. Evaluasi dilakukan terhadap model tersebut dengan mengukur Root Mean Square Error (RMSE) dan Kesalahan Relatif (Relative Error). Hasil evaluasi menunjukkan bahwa nilai RMSE untuk kategori harga adalah 1512.262 dan kesalahan relatif sebesar 9.13%.

Kata Kunci—prediksi, harga, dataset, regresi linear berganda, model regresi

Abstract— *The lack of information about used car prices can hinder people from engaging in vehicle transactions because they may not understand the relevant market prices. Therefore, having accurate and reliable information is crucial for determining estimated prices as a reference. In addressing this issue, the use of multiple linear regression algorithms can assist in predicting used car prices. In this study, a multiple linear regression algorithm is utilized to predict used car prices. The data used to build the prediction model are sourced from Kaggle repositories. The testing process of the model involves splitting the data into training data (90%) and testing data (10%). Following the training process, the prediction model is obtained with an average price value of 1512.262. Evaluation is performed on the model by measuring the Root Mean Square Error (RMSE) and Relative Error. The evaluation results indicate that the RMSE value for the price category is 1512.262, with a relative error of 9.13%.*

Keywords—prediction, price, dataset, multiple linear regression, regression mode.

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Nama Penulis, Erna Daniati
Departemen Penulis, Sistem Informasi
Institusi Penulis, Universitas Nusantara PGRI Kediri
Email: ernadaniati@unpkediri.ac.id
ID Orcid: [<https://orcid.org/0009-0008-9471-4421>]
Handphone: 081335242202

I. PENDAHULUAN

Saat ini, industri otomotif sedang mengalami perkembangan pesat dengan persaingan yang semakin ketat antara perusahaan dan penggunaan teknologi yang semakin maju. Banyaknya kendaraan baru yang diproduksi oleh pabrikan mengikuti tren yang berkembang juga memberikan dampak signifikan. Kendaraan bermotor memainkan peran penting dalam memfasilitasi mobilitas dan pertumbuhan ekonomi masyarakat.

Karena kondisi ini, perusahaan otomotif harus berpikir keras untuk tetap eksis. Sebagai hasilnya, peningkatan produksi kendaraan baru berdampak pada banyaknya mobil bekas yang tersedia di pasaran. Dalam proses jual-beli mobil bekas, harga mobil cenderung mengalami penurunan secara berkala. Ini berarti mobil bekas dengan tahun yang lebih tua dapat memiliki harga jual yang lebih rendah dibandingkan dengan mobil bekas yang lebih baru, terlepas dari kondisi yang sama.

Untuk mengatasi tantangan ini, penelitian telah dilakukan untuk mengembangkan model prediksi harga mobil bekas yang akurat menggunakan metode regresi linier berganda. Dengan model ini, diharapkan dapat diidentifikasi variabel yang mempengaruhi harga mobil dan memprediksi harga berdasarkan karakteristik dan kondisi pasar yang relevan. Penelitian ini berbeda dengan penelitian sebelumnya karena menggunakan algoritma regresi linier dan mengimplementasikannya dalam platform online. Selain itu, prediksi harga mobil bekas melibatkan penggunaan platform RapidMiner untuk melakukan analisis data dan membangun model prediksi harga mobil bekas dengan berbagai algoritma pembelajaran mesin yang tersedia. Penelitian ini menggunakan metode regresi linier berganda untuk analisis data. Tujuan dari analisis data ini adalah untuk memprediksi harga mobil bekas [1]. Metode ini memungkinkan kita mempelajari beberapa faktor sebelumnya untuk meramalkan kejadian di masa depan. Dengan cara ini, kita bisa lebih siap menghadapi apa yang akan terjadi nantinya [2].

Metode yang tepat guna memprediksi harga mobil bekas untuk meningkatkan efisiensi operasional dan memaksimalkan keuntungan yaitu Regresi Linear [3]. Penelitian ini bertujuan untuk menerapkan regresi linear menggunakan RapidMiner sebagai alat bantu analisis dalam memprediksi harga mobil bekas. Dengan memanfaatkan RapidMiner, diharapkan pelaku bisnis dapat mengoptimalkan proses penentuan harga dan mengambil keputusan yang lebih tepat berdasarkan prediksi yang akurat. Meskipun sektor penjualan mobil bekas sangat penting bagi ekonomi nasional, mereka sering menghadapi banyak tantangan, terutama dalam menentukan harga yang sesuai dan memprediksi permintaan. Kesalahan dalam penentuan harga dapat

menyebabkan harga yang terlalu tinggi atau terlalu rendah, yang berdampak pada biaya operasional dan kepuasan pelanggan. Oleh karena itu, pengabdian kepada masyarakat melalui penerapan teknologi analisis data sangat penting untuk membantu pelaku bisnis mengatasi masalah ini. Beberapa masalah utama dalam penjualan mobil bekas adalah harga yang sulit diprediksi dan kurangnya efisiensi dalam manajemen penawaran dan permintaan. Risiko menentukan harga yang tidak sesuai meningkat tanpa prediksi yang akurat, yang dapat merugikan bisnis. Dalam manajemen bisnis penjualan mobil bekas, teknologi analisis data memiliki potensi besar untuk meningkatkan profitabilitas dan efisiensi operasional, meskipun teknologi analisis data dalam sektor ini masih sangat baru. Pelaku bisnis dapat mengelola penawaran dan permintaan dengan lebih baik, mengurangi biaya operasional, dan meningkatkan kepuasan pelanggan dengan memprediksi harga secara akurat. Dengan tujuan memudahkan penentuan harga produk, pelaku bisnis akan lebih mudah memenuhi permintaan pasar [4].

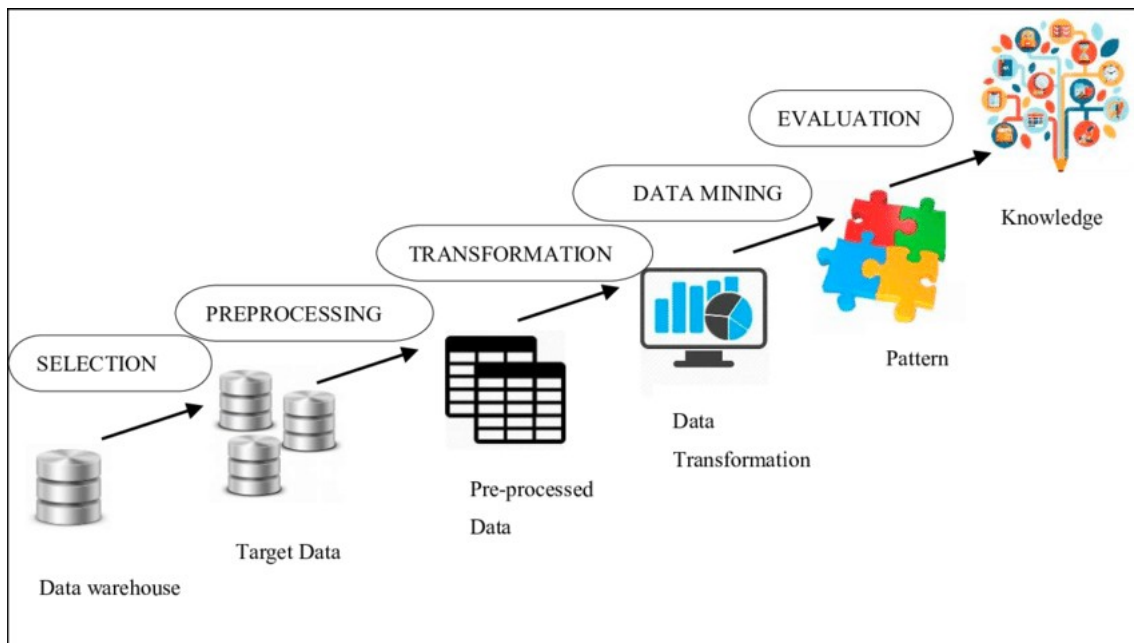
Semakin mendekati nilai RMSE yang mendekati 0, semakin tinggi akurasi prediksinya [5]. RapidMiner adalah platform analisis data yang memungkinkan pengguna melakukan pemodelan prediktif tanpa memerlukan keahlian pemrograman yang mendalam. RapidMiner dikembangkan dengan pendekatan open core, yang menggabungkan elemen-elemen open source dengan fitur-fitur tambahan yang tersedia dalam versi berbayar [6]. Data Mining adalah proses mengidentifikasi korelasi, pola, dan tren baru yang berarti dengan memeriksa sejumlah besar data yang tersimpan dalam repositori, menggunakan teknologi pengenalan pola, serta metode statistik dan matematika [7]. Data Mining merupakan proses menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang berguna serta pengetahuan yang terkait dari berbagai database besar [8]. Data mining adalah rangkaian langkah untuk menemukan nilai tambah dari sekumpulan data dengan mengungkap pengetahuan yang sebelumnya tidak terdeteksi secara manual [9]. Tujuan utama data mining adalah mengungkap, menemukan, atau mengeksplorasi pengetahuan dari data atau informasi yang tersedia [10].

Penelitian yang akan dilakukan ini menerapkan Algoritma Regresi Linear Berganda, karena metode ini sesuai dengan studi kasus yang sedang diteliti. Data yang digunakan dalam penelitian ini berasal dari Kaggle. Sistem prediksi ini bertujuan untuk memperkirakan harga mobil bekas menggunakan Algoritma Regresi Linear Berganda. Penelitian oleh Wati, Sebayang, & Sitepu (2013) mengenai peramalan jumlah produksi menunjukkan bahwa regresi linier berganda lebih unggul dibandingkan dengan metode fuzzy, dengan rata-rata kesalahan relatif regresi linier berganda sebesar 9,383% yang lebih rendah dibandingkan dengan nilai rata-rata kesalahan relatif fuzzy sebesar 20,748% [11]. Abdulah Sajad dan Nurmalitasari juga

menggunakan metode regresi linear dalam Analisis Faktor-Faktor yang Mempengaruhi Harga Mobil Bekas dan menghasilkan keakuratan prediksi yang cukup tinggi dengan nilai error sebesar 0,17 [12]. Dari berbagai penelitian tersebut dapat disimpulkan bahwa metode Regresi Linear Berganda efektif untuk memprediksi sesuatu di masa depan dengan menggunakan lebih dari dua faktor atau variabel yang mempengaruhi hasil yang diprediksi.

II. METODE

Metodologi ini menguraikan lima langkah yang diperlukan untuk memprediksi jumlah peminat mata kuliah pilihan menggunakan regresi linier. Tahapan metodologi penelitian ini adalah sebagai berikut:



Gambar 1. Metode knowledge discovery in databases (KDD)

Berikut adalah langkah-langkah yang perlu dilakukan dalam penelitian berdasarkan tahapan knowledge discovery in databases:

1. Seleksi Data: Ini melibatkan pengambilan, pemilihan, dan pengecekan informasi untuk proses selanjutnya.
2. Pembersihan Data: Langkah ini bertujuan untuk menghilangkan data yang tidak memiliki nilai (null), yang dapat mempengaruhi kinerja sistem data mining dengan mengurangi jumlah dan kompleksitas data yang akan ditangani.

3. Transformasi Data: Tujuannya adalah untuk mempermudah koordinasi data yang diproses oleh algoritma dan alat yang digunakan dalam penelitian. Data transformation adalah Transformasi data yang telah dipilih bertujuan agar data tersebut sesuai untuk proses penambangan data. Pada tahap transformasi ini, data yang awalnya bertipe polinomial diubah menjadi tipe numerik agar data tersebut dapat diolah berdasarkan jarak. Untuk melakukan perubahan dari tipe polinomial menjadi numerik, digunakan operator Nominal to Numerical. Operator ini mengonversi nilai-nilai kategorikal (polinomial) menjadi nilai-nilai numerik yang dapat diukur dan dianalisis dengan lebih efektif dalam proses data mining [13].
4. Penambangan Data: Proses ini melibatkan pengolahan data berdasarkan algoritma yang konsisten dengan alur data mining.
5. Evaluasi: Tahap ini digunakan untuk mengevaluasi hasil dari algoritma Regresi Linear untuk menarik kesimpulan dari proses data mining. Kinerja diukur dengan menggunakan nilai RMSE (Root Mean Squared Error).

III. HASIL DAN PEMBAHASAN

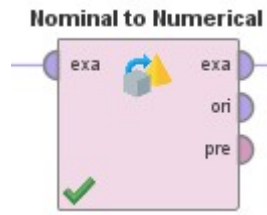
Dalam penelitian ini, dataset yang digunakan terdiri dari informasi tentang harga mobil bekas beserta beberapa atribut, dengan total 6738 data. Data ini diunduh dari situs Kaggle, dengan sumber yang berasal dari (<https://www.kaggle.com/datasets/adityadesai13/used-car-dataset-ford-and-mercedes/data?select=toyota.csv>). Dataset sampel dapat dilihat pada Tabel 1.

Tabel 1. Dataset Toyota

NO	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
1	GT86	2016	16000	Manual	24089	Petrol	265	36.200	2
2	GT86	2017	15995	Manual	18615	Petrol	145	36.200	2
3	GT86	2015	13998	Manual	27469	Petrol	265	36.200	2
4	Urban Cruiser	2012	4995	Manual	46000	Diesel	125	57.600	1.300
5	Urban Cruiser	2011	3995	Manual	60700	Petrol	125	50.400	1.400
6	Urban Cruiser	2011	4495	Manual	45128	Petrol	125	50.400	1.400

Data Selection

Mengubah tipe data nominal menjadi numerik. Atribut yang berisi data nominal diubah menjadi data numerical. Gambar 2. Nominal to Numeric



Gambar 2. Nominal to Numeric

Data Preprocessing

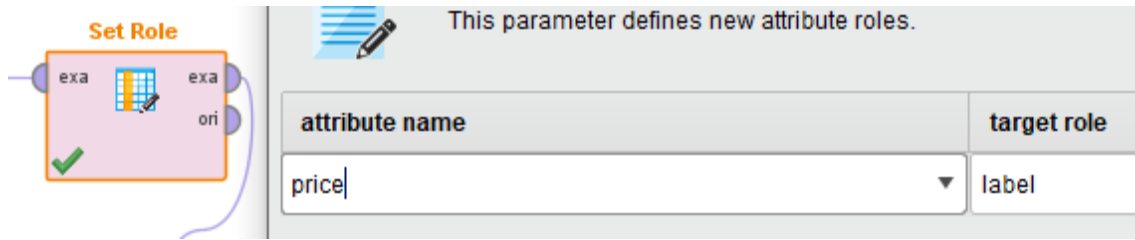
Pada tahap ini, dilakukan proses pra-pemrosesan data yang meliputi pembersihan data untuk mengatasi nilai yang hilang atau kosong, sebagaimana diperlihatkan dalam Gambar 3.

▼ model	Nominal	0	Least Verso-S (3)	Most Yaris (2122)	Values Yaris (2122), Aygo (1961), ...[16 more]
▼ year	Integer	0	Min 1998	Max 2020	Average 2016.748
▼ price	Integer	0	Min 850	Max 59995	Average 12522.391
▼ transmission	Nominal	0	Least Other (1)	Most Manual (3826)	Values Manual (3826), Automatic (2657), ...[2 more]
▼ mileage	Integer	0	Min 2	Max 174419	Average 22857.414
▼ fuelType	Nominal	0	Least Other (105)	Most Petrol (4087)	Values Petrol (4087), Hybrid (2043), ...[2 more]
▼ tax	Integer	0	Min 0	Max 565	Average 94.697
▼ mpg	Real	0	Min 2.800	Max 235	Average 63.042

Gambar 3. Data Missing Value

Transformation

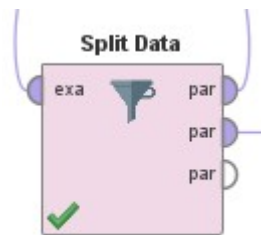
Transformasi data ke dalam format yang sesuai untuk diproses oleh algoritma data mining dapat dilihat dalam Gambar 4.



Gambar 4. Pemilihan Label

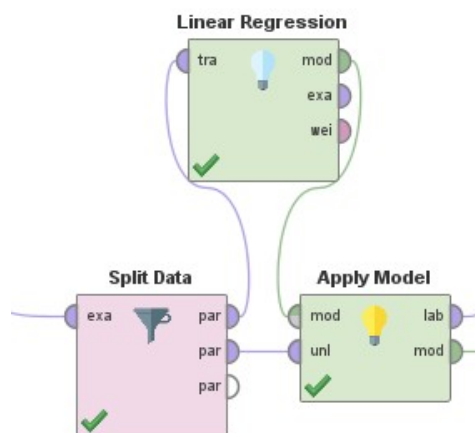
Data Mining

Dalam penelitian ini, proses data mining yang diterapkan adalah prediksi menggunakan regresi linear, di mana evaluasi hasil prediksi diukur dengan nilai RMSE. Proses ini dilakukan menggunakan aplikasi RapidMiner. Data training digunakan untuk membangun model regresi linear berganda, sementara data testing digunakan untuk menguji kinerja model yang telah dilatih. Pembagian data training dan testing dapat dilihat dalam Gambar 5.



Gambar 5. Operator Split Data

Setelah itu, dilakukan pemodelan dengan model yang ditunjukkan dalam Gambar 6.



Gambar 6. Proses Pemodelan

Evaluasi

Pengujian koefisien korelasi berguna untuk mengevaluasi sejauh mana variabel independen memengaruhi variabel dependen dalam rangka menilai akurasi model regresi. Proses ini dapat disajikan dalam Tabel 2.

RMSE	1512.262
Squared_error	2286936.659
Squared_correlation	0.931

Tabel 2. Evaluasi Performance

Untuk mengevaluasi hasil prediksi harga mobil, digunakan perhitungan RMSE. Semakin mendekati nilai terkecil, semakin baik hasil prediksi tersebut. Evaluasi hasil dapat dilihat dalam Gambar 7.

root_mean_squared_error

```
root_mean_squared_error: 1512.262 +/- 0.000
```

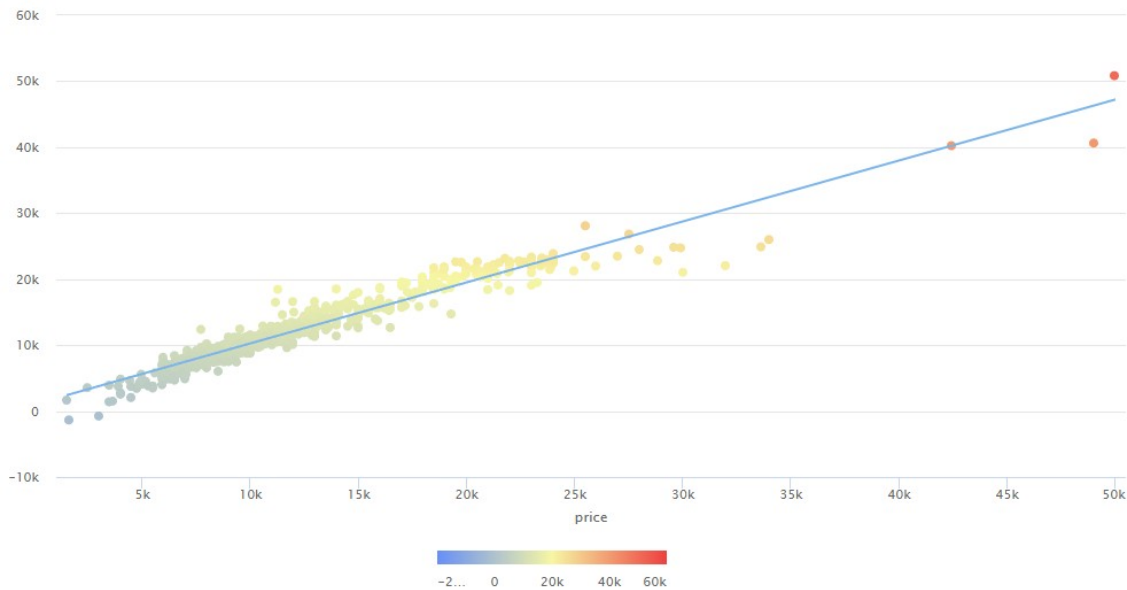
relative_error

```
relative_error: 9.13% +/- 11.75%
```

Gambar 7. Hasil Nilai RMSE dan Relative Error

Knowledge

Hasil prediksi harga mobil bekas menggunakan algoritma regresi linear berganda dengan tahapan proses Knowledge Discovery in Databases (KDD) telah berhasil ditampilkan. Pada kategori prediksi harga, plot scatter bubble menampilkan hasil prediksi dari model regresi linear. Hasil prediksi menunjukkan variasi nilai harga yang beragam, seperti yang terlihat pada Gambar 8.



Gambar 8. Plot Prediksi Harga

Pembahasan

Nilai evaluasi yang diperoleh dari perhitungan prediksi menggunakan RMSE (Root Mean Squared Error) adalah 1512.262. Pada tahap evaluasi performa, penggunaan RMSE menunjukkan tingkat kesalahan sebesar 1512.262.

Hasil prediksi menggunakan algoritma regresi linear menghasilkan model dengan nilai RMSE 1512.262. Model prediksi ini akan dievaluasi dengan dua indikator penilaian, yaitu RMSE dan Relative Error. Hasil yang diperoleh dari model tersebut menunjukkan bahwa pada kategori harga, nilai RMSE adalah 1512.262 dan nilai Relative Error adalah 9.13%. Pendekatan ini mengukur kesalahan dengan menghitung akar kuadrat dari selisih antara nilai sebenarnya dan nilai prediksi.

Menurut hasil penelitian (Hamdanah & Fitriyah, 2021) nilai RMSE yang sangat baik adalah yang mendekati 0 [14], lalu hasil evaluasi Relative Error $\leq 25\%$, dianggap sebagai tingkat akurasi yang baik. Jika nilainya $> 25\%$ tetapi $\leq 50\%$, mungkin masih dapat dipertimbangkan [15].

IV. KESIMPULAN

Dari penelitian yang telah dilakukan menggunakan algoritma Regresi Linear untuk memprediksi harga mobil, diperoleh hasil prediksi yang dievaluasi dengan mengukur akurasi

error menggunakan RMSE dan Relative Error. Hasil evaluasi menunjukkan nilai RMSE sebesar 1512.262 +/- 0.000 dan Relative Error sebesar 9.13% +/- 11.75%. Berdasarkan hasil ini, prediksi harga mobil dianggap baik dan cukup akurat, menunjukkan bahwa metode prediksi ini dapat diandalkan.

DAFTAR PUSTAKA

- [1] M. Z. Sasongko, S. Sucipto, and E. Daniati, "Perencanaan Sistem Manajemen Evaluasi Belajar Mahasiswa Berbasis Elearning Dengan Pendekatan Berorientasi Objek," *Respati*, vol. 13, no. 2, pp. 10–19, 2018, doi: 10.35842/jtir.v13i2.228.
- [2] A. C. Pradikdo and A. Ristyawan, "Model Klasifikasi Abstrak Skripsi Menggunakan Text Mining Untuk Pengkategorian Skripsi Sesuai Bidang Kajian," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 9, no. 2, pp. 1091–1098, 2018, doi: 10.24176/simet.v9i2.2513.
- [3] A. Forecasting Penjualan Dan Persediaan Produk Usaha Mikro, K. Dan Menengah, R. Jumardi, S. Handani Widiastuti, and S. Tinggi Teknologi Bontang, "Article Info ABSTRAK."
- [4] H. Jaya, R. Gunawan, and R. Kustini, "Penerapan Data Mining Untuk Memprediksi Target Produksi Berdasarkan Tingkat Penjualan Dan Banyaknya Pemesanan Produk Pada Pt. Neo National Menggunakan Metode Regresi Linier Berganda," vol. 18, no. SAINTIKOM, pp. 219–227, 2019.
- [5] D. S. Purnia, H. Muhajir, M. F. Adiwisastra, and D. Supriadi, "Pengukuran Kesenjangan Digital Menggunakan Metode Deskriptif Berbasis Website," *Jurnal Sains dan Manajemen*, vol. 8, no. 2, 2020.
- [6] R. Nofitri and N. Irawati, "Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, pp. 199–204, Jul. 2019, doi: 10.33330/jurteksi.v5i2.365.
- [7] D. Setiawan, N. Surojudin, and W. Hadikristanto, "Tahun 2022 Call for papers dan Seminar Nasional Sains dan Teknologi Ke-1 2022 Fakultas Teknik, Universitas Pelita Bangsa," vol. 1, no. 1, 2022.
- [8] K. Tingkat *et al.*, "Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi."
- [9] R. Saepul Rohman, R. Amegia Saputra, and D. Arif Firmansaha, "Komparasi Algoritma C4.5 Berbasis Pso Dan Ga Untuk Diagnosa Penyakit Stroke," 2020. [Online]. Available: <https://www.kaggle.com/asaumya/healthcare-dataset->
- [10] A. Supriyadi Sunge and A. Turmudi Zy, "Analisis Prediksi Penjualan Dengan Metode Regresi Linear Di Pt. Eagle Industry Indonesia," 2023.

- [11] S. E. Wati, D. Sebayang, and R. Sitepu, “Perbandingan Metode Fuzzzy Dengan Regresi Linier Berganda Dalam Peramalan Jumlah Produksi,” *Saintia Matematika*, vol. 1, no. 3, pp. 273–284, 2013.
- [12] A. Sajad and N. Nurmalitasari, “Analisis Faktor–Faktor Yang Mempengaruhi Harga Mobil Bekas Menggunakan Metode Regresi Linier,” *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis*, pp. 266–270, 2023.
- [13] A. Muhamad Andika, N. Suarna, and R. D. Dana, “Analisa Dataset Asosiasi Penjualan Menggunakan Metode FP-Growth,” *Jurnal Teknologi Ilmu Komputer*, vol. 2, no. 1, pp. 80–88, 2023, doi: 10.56854/jtik.v2i1.108.
- [14] F. H. Hamdanah and D. Fitriana, “Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikro, Kecil, dan Menengah,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 10, no. 1, p. 23, Apr. 2021, doi: 10.23887/janapati.v10i1.31035.
- [15] Y. A. Ramadhan, A. Faqih, and G. Dwilestari, “Jurnal Informatika Terpadu Prediksi Penjualan Handphone Di Toko X Menggunakan Algoritma Regresi Linear,” *Jurnal Informatika Terpadu*, vol. 9, no. 1, pp. 40–44, 2023, [Online]. Available: <https://journal.nurulfikri.ac.id/index.php/JIT>