

IMPLEMENTASI XGBOOST DAN LOGIKA REVERSE CALCULATION PADA SISTEM ESTIMASI HARGA BELI MOBIL BEKAS BERBASIS WEB

Ida Bagus Aditya Cahya Wiraguna✉, Putu Indah Ciptayani,

I Putu Bagus Arya Pradnyana, Ni Gusti Ayu Putu Harry Saptarini

Teknologi Rekayasa Perangkat Lunak, Jurusan Teknologi Informasi, Politeknik Negeri Bali, Badung, Indonesia

Email: guswiraaditya@gmail.com

DOI: <https://doi.org/10.46880/jmika.Vol10No1.pp357-365>

ABSTRACT

The movement of used car prices is highly dynamic and subjective, making the manual estimation process prone to bias and financial risk. This study aims to develop a web-based used car purchase price estimation system using the Extreme Gradient Boosting (XGBoost) algorithm combined with a reverse calculation logic. The dataset was obtained from secondary market data and primary showroom transaction records, totaling 12,324 clean data after passing the Grouped-IQR outlier filtering process. The XGBoost model was optimized using Grid Search and validated through 10-Fold Cross-Validation. The results showed that the optimal model configuration achieved a Mean Absolute Percentage Error (MAPE) of 11.23%, a Root Mean Squared Error (RMSE) of Rp 54,779,437, and a Coefficient of Determination (R²) of 0.8386. This performance indicates a highly accurate forecasting capability. The model was successfully integrated into a Laravel-based web application via a Python REST API, allowing users to obtain fair market price predictions and maximum purchase bids to improve the efficiency and objectivity of decision-making.

Keyword: XGBoost, Price Estimation, Used Cars, Machine Learning, Web-Based System.

ABSTRAK

Pergerakan harga mobil bekas sangat dinamis dan subjektif, sehingga proses penaksiran manual rentan terhadap bias dan risiko finansial. Penelitian ini bertujuan mengembangkan sistem estimasi harga beli mobil bekas berbasis web menggunakan algoritma Extreme Gradient Boosting (XGBoost) yang dipadukan dengan logika kalkulasi mundur (reverse calculation). Dataset diperoleh dari data sekunder pasar otomotif dan catatan transaksi primer showroom, menghasilkan 12.324 data bersih setelah melalui proses penyaringan outlier Grouped-IQR. Model XGBoost dioptimasi menggunakan Grid Search dan divalidasi secara ketat melalui 10-Fold Cross-Validation. Hasil pengujian menunjukkan bahwa konfigurasi model optimal mencapai nilai Mean Absolute Percentage Error (MAPE) sebesar 11,23%, Root Mean Squared Error (RMSE) sebesar Rp 54.779.437, dan Koefisien Determinasi (R²) sebesar 0.8386. Performa ini mengindikasikan tingkat akurasi prediksi yang sangat baik dan melampaui algoritma konvensional. Model berhasil diintegrasikan ke dalam aplikasi web berbasis Laravel melalui layanan REST API Python, memungkinkan pengguna mendapatkan prediksi harga pasar wajar dan batas harga tawaran beli maksimal guna meningkatkan efisiensi dan objektivitas pengambilan keputusan.

Kata Kunci: XGBoost, Estimasi Harga, Mobil Bekas, Machine Learning, Sistem Berbasis Web.

PENDAHULUAN

Sektor otomotif di Indonesia merupakan salah satu pilar strategis yang menopang pertumbuhan ekonomi nasional. Merujuk pada statistik pasar tahun 2024, ekosistem perdagangan mobil bekas menunjukkan volume yang besar dengan proyeksi penjualan menyentuh angka 1,8 juta unit. Angka ini secara signifikan melampaui serapan pasar mobil baru yang berada pada level 865.723 unit (Arradian, 2025). Tingginya perputaran transaksi ini merepresentasikan peluang bisnis yang menjanjikan, namun di saat bersamaan memunculkan tantangan manajerial yang kompleks. Kunci profitabilitas usaha pada sektor ini

sangat bergantung pada efisiensi manajemen persediaan, terutama pada fase akuisisi unit. Secara mendasar, margin laba perusahaan terbentuk dari selisih antara harga wajar kendaraan di pasaran dengan total biaya pengadaan di awal transaksi.

Pada praktiknya, proses pengadaan kendaraan masih menghadapi berbagai kendala teknis, sebagaimana studi kasus pada CV Guna Abadi. Mekanisme penaksiran harga beli saat ini masih bertumpu pada intuisi negosiasi dan inspeksi fisik yang dilakukan secara manual oleh staf penaksir (*appraiser*). Pendekatan konvensional ini memunculkan rentang subjektivitas yang tinggi, sehingga penetapan batas



maksimal harga beli seringkali tidak konsisten. Di samping itu, ketiadaan sistem referensi harga seketika (*real-time*) membuat proses pencarian data dari berbagai sumber eksternal menjadi kurang efisien. Lambatnya akses terhadap fluktuasi harga pasar dan standar biaya restorasi pada akhirnya menghambat kecepatan pengambilan keputusan, menurunkan produktivitas operasional, dan memperbesar risiko kerugian finansial akibat pembelian inventaris di atas harga kewajaran (*overpriced*).

Sebagai solusi atas permasalahan tersebut, pendekatan komputasi berbasis kecerdasan buatan (*Machine Learning*) diterapkan untuk membangun sistem estimasi harga yang terukur dan objektif. Secara global, penerapan algoritma *Machine Learning* telah terbukti krusial dalam industri bisnis otomotif modern untuk memfasilitasi penetapan harga secara dinamis dan akurat, mengingat fluktuasi harga aset di pasar sekunder sangat rentan terhadap berbagai faktor multidimensi (Yang et al., 2024). Pengembangan model prediktif dalam riset ini didukung oleh basis data yang memadukan data primer dari riwayat transaksi internal CV Guna Abadi dengan data sekunder hasil ekstraksi (*web scraping*) dari *marketplace* otomotif nasional, seperti OLX dan Oto. Penggabungan data dari kedua sumber ini penting untuk menangkap pola nyata penyusutan nilai (depresiasi) aset kendaraan di pasar.

Berdasarkan penelusuran literatur, implementasi algoritma regresi untuk taksiran harga aset telah banyak diuji. Meskipun metode *K-Nearest Neighbor* (KNN) dinilai cukup aplikatif pada studi terdahulu (Haya & Nudin, 2024), algoritma ini terbukti memiliki kelemahan struktural terhadap anomali data serta sangat bergantung pada kalibrasi parameter jarak (Ghaisani & Baita, 2025). Sebaliknya, evaluasi algoritma menunjukkan bahwa metode *tree-ensemble*, khususnya *Gradient Boosting*, memiliki performa yang lebih unggul dibandingkan Regresi Linier maupun *Random Forest* dalam memproses data otomotif yang berwujud tabular (Idris & Pontooyo, 2025; Winata, 2024). Merujuk pada temuan tersebut, varian *Extreme Gradient Boosting* (XGBoost) dipilih karena keandalannya dalam komputasi serta kemampuannya mereduksi risiko *overfitting* pada karakteristik data pasar yang sangat volatil (Eka Adinagoro et al., 2025).

Kebaruan (*novelty*) dari penelitian ini tidak hanya terletak pada penggunaan algoritma XGBoost, melainkan pada inovasi pra-pemrosesan data berbasis domain (*domain-aware preprocessing*) dan arsitektur integrasinya. Keandalan model ditingkatkan melalui penyaringan data pencilan spesifik per segmen merek menggunakan *Grouped Interquartile Range* (Grouped-IQR) dan penerapan *Target Encoding* untuk

memproses fitur berkardinalitas tinggi secara efisien. Validitas sistem selanjutnya diuji secara ketat menggunakan skema *10-Fold Cross-Validation*. Pada tahap akhir, hasil estimasi model AI ditransmisikan secara *real-time* ke dalam antarmuka web Laravel. Melalui logika kalkulasi mundur (*reverse calculation*), sistem secara otomatis mendeduksi target margin dan estimasi biaya perbaikan dari harga pasar untuk menghasilkan rekomendasi batas harga beli. Oleh karena itu, penelitian ini bertujuan merancang dan mengevaluasi sistem estimasi harga pengadaan mobil bekas berbasis web yang andal, menggunakan metrik *Coefficient of Determination* (R^2), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) sebagai instrumen validasi keputusan finansial.

KAJIAN LITERATUR

Pengembangan model komputasi untuk prediksi harga kendaraan telah melalui berbagai evolusi pendekatan algoritmik. Pada fase awal, pendekatan heuristik berbasis kedekatan jarak seperti *K-Nearest Neighbor* (KNN) banyak diadopsi untuk mengestimasi nilai pasar (Haya & Nudin, 2024). Meskipun metode regresinya cukup intuitif, evaluasi kinerja KNN menunjukkan kerentanan yang signifikan terhadap data pencilan (*outlier*). Selain itu, ketergantungan metode ini pada proses *trial-and-error* untuk menentukan parameter jarak yang optimal menjadikannya kurang efisien (Ghaisani & Baita, 2025). Mengingat pergerakan harga mobil bekas yang sangat fluktuatif, diperlukan algoritma yang lebih adaptif dan stabil.

Sebagai alternatif pemecahan masalah regresi pada data tabular, algoritma berbasis *tree-ensemble* terbukti memberikan performa yang lebih unggul. Studi komparasi yang mengevaluasi algoritma Regresi Linier, *Random Forest*, dan *Gradient Boosting* pada ekosistem otomotif menyimpulkan bahwa *Gradient Boosting* menghasilkan tingkat akurasi prediksi tertinggi dengan koefisien determinasi mencapai 0,96 (Idris & Pontooyo, 2025). Temuan empiris ini konsisten dengan riset global berskala besar yang membuktikan bahwa integrasi algoritma berbasis pohon seperti XGBoost dengan teknik *feature engineering* tingkat lanjut secara konklusif mampu melampaui kinerja regresi konvensional dalam memodelkan kompleksitas heterogenitas harga mobil bekas (Fayyaz et al., 2025). Temuan ini diperkuat oleh penelitian lain yang membuktikan bahwa varian *Extreme Gradient Boosting* (XGBoost) mampu mencapai kinerja terbaik pada dataset dengan kardinalitas tinggi berkat mekanisme regularisasinya yang secara efektif mencegah *overfitting* (Eka Adinagoro et al., 2025;

Syafarina et al., 2026). Dalam konteks otomotif, algoritma berbasis pohon keputusan ini sangat relevan untuk menangkap pola depresiasi kendaraan yang bersifat *non-linear* seiring bertambahnya umur dan jarak tempuh (Safitri et al., 2024).

Meskipun algoritma XGBoost dikenal tangguh, performa prediksi pada data sekunder sangat bergantung pada kualitas rekayasa fitur (*feature engineering*). Mayoritas penelitian sebelumnya seringkali menghadapi masalah komputasi akibat penggunaan *One-Hot Encoding* secara menyeluruh. Pendekatan ini memicu masalah ledakan dimensi (*curse of dimensionality*) pada variabel kategorikal dengan jumlah kelas unik yang tinggi, seperti atribut "Merek" dan "Model" (Bolikulov et al., 2024). Selain itu, penanganan data pencilan harga yang dilakukan secara global (*Global IQR*) seringkali keliru menghapus data transaksi mobil segmen premium yang sewajarnya bernilai tinggi. Sebagai solusi, penelitian ini menerapkan pendekatan *domain-aware*, yang meliputi *Target Encoding* untuk representasi kategorikal yang efisien, serta *Grouped IQR* untuk penyaringan anomali harga berbasis segmen merek kendaraan.

Celah penelitian (*research gap*) lain yang diidentifikasi dari literatur terdahulu adalah terbatasnya fokus luaran dari sistem yang dikembangkan. Sebagian besar model kecerdasan buatan hanya difungsikan untuk memprediksi estimasi harga jual wajar bagi konsumen akhir (Ghaisani & Baita, 2025; Winata, 2024). Pendekatan sistematis yang mengonversi hasil prediksi tersebut menjadi instrumen taktis pengadaan unit (*procurement*) bagi entitas bisnis belum banyak dikembangkan. Untuk menjembatani hal ini, penelitian ini mengimplementasikan konsep penetapan harga *mark-up* (Ariyono & Manafe, 2022) melalui skema kalkulasi terbalik (*Reverse Calculation*). Melalui integrasi *Application Programming Interface* (API) antara *backend* Laravel dan layanan mikro Python, target laba operasional dan estimasi biaya perbaikan dideduksi secara otomatis dari prediksi harga XGBoost. Sinergi antara komputasi *Data Science* dan logika bisnis empiris ini memastikan sistem tidak hanya berfungsi sebagai alat prediksi, melainkan instrumen validasi keputusan finansial yang komprehensif.

METODE PENELITIAN

Penelitian ini dirancang menggunakan kerangka riset kuantitatif terapan (*applied research*) yang ditujukan untuk memecahkan problematika empiris terkait bias estimasi harga di ekosistem bisnis otomotif, dengan memusatkan studi kasus pada operasional CV

Guna Abadi. Lingkup penelitian merentang dari perumusan model prediktif berbasis Machine Learning hingga integrasi teknisnya ke dalam arsitektur sistem informasi berbasis web.

Pengumpulan Data dan Identifikasi Variabel

Basis data yang digunakan dalam penelitian ini merupakan gabungan dari dataset sekunder dan primer dengan total keseluruhan mencapai 13.362 baris data. Dataset sekunder sejumlah 13.261 data diekstraksi menggunakan teknik web *scraping* dari platform *marketplace* otomotif nasional (OLX dan Oto). Proses ekstraksi dilakukan secara dinamis menggunakan *library* Selenium WebDriver untuk merender elemen *lazy-loading* berbasis JavaScript, yang dipadukan dengan teknik *DOM parsing* dan *Regular Expression* (RegEx) guna mengurai atribut spesifik kendaraan dari teks deskripsi yang tidak terstruktur. Sementara itu, dataset primer sejumlah 101 data dikompilasi dari riwayat rekapitulasi transaksi internal CV Guna Abadi. Data tersebut memuat variabel prediktor (*independen*) yang terdiri dari atribut fisik kendaraan, meliputi: Merek, Model, Tipe Bodi, Kapasitas Mesin, Tahun Pembuatan, Transmisi, Warna, Wilayah, Jarak Tempuh, dan Bahan Bakar. Variabel target (*dependen*) yang digunakan untuk melatih model adalah Harga Jual aktual di pasar. Sementara itu, untuk keperluan kalkulasi bisnis, sistem juga mengakomodasi variabel operasional yang diinputkan pengguna secara dinamis, yaitu Harga Penawaran Penjual, Target Margin Keuntungan, dan Estimasi Biaya Perbaikan.

Pra-pemrosesan Data (Data preprocessing)

Untuk memastikan kualitas pembelajaran algoritma, data mentah diproses melalui tahapan otomatis:

1. Pembersihan Data (*Data Cleansing*)

Menghapus data duplikat dan anomali nilai berdasarkan logika bisnis (misalnya harga bernilai negatif atau tahun pembuatan yang melampaui tahun berjalan). Selanjutnya, penanganan nilai pencilan (*outlier*) pada variabel numerik seperti Harga dan Jarak Tempuh dilakukan menggunakan metode *Grouped Interquartile Range* (IQR) dengan batas pengali (*multiplier*) 1.5. Perhitungan batas bawah dan batas atas IQR tidak dilakukan secara global, melainkan dihitung secara spesifik untuk masing-masing kelompok Merek mobil (*Grouped-IQR*). Pendekatan berbasis domain (*domain-aware*) ini sangat penting untuk mempertahankan keaslian sebaran harga pada segmentasi mobil mewah (*premium market*) sekaligus secara efektif membuang entri data yang benar-benar merupakan

anomali atau kesalahan input di dalam kelas merek tersebut. Ketidakeengkapan data (*missing values*) kemudian ditangani menggunakan teknik imputasi Median untuk atribut numerik dan nilai terbanyak (*most frequent*) untuk atribut kategorikal.

2. Rekayasa Fitur (*Feature Engineering*)

Atribut *Tahun Pembuatan* dikonversi menjadi *Umur Kendaraan* untuk menangkap pola penyusutan harga aset yang lebih representatif. Selain itu, diciptakan fitur turunan berupa rasio Kilometer per Tahun (*Km_per_Tahun*) untuk mengukur intensitas pemakaian secara historis, serta merangkai atribut Merek, Model, dan Varian menjadi satu identitas kolom baru (*Kendaraan Lengkap*) guna menangkap nilai jual unik dari varian trim spesifik kendaraan tersebut.

3. Pengodean Kategorikal Hibrida (*Hybrid Categorical Encoding*)

Konversi data teks ke numerik diterapkan secara selektif melalui *pipeline* transformasi bersyarat. *One-Hot Encoding* digunakan untuk fitur kategorikal dengan kardinalitas rendah (Warna, Bahan Bakar, Tipe Bodi). Target Encoding (rata-rata terbobot variabel dependen) diterapkan untuk mereduksi dan menangkap informasi dari fitur-fitur berkardinalitas tinggi secara efisien (Merek, Model, Varian, Wilayah, Kendaraan Lengkap) tanpa menyebabkan ledakan dimensi matriks (*dimensionality explosion*). Secara teoretis dan empiris, penerapan regularisasi Target Encoding di dalam *pipeline* pembelajaran mesin telah diakui sebagai metode *state-of-the-art* yang terbukti tangguh dalam mencegah kebocoran data (*data leakage*) sekaligus secara konsisten melampaui efektivitas pengodean konvensional (Pargent et al., 2022). Sementara itu, *Ordinal Encoding* digunakan khusus pada fitur kategorikal yang memiliki tingkat atau kelas biner (Transmisi).

4. Transformasi Distribusi Target

Distribusi variabel harga yang fluktuatif dan condong (*skewed*) distabilkan menggunakan transformasi logaritma natural, $y' = \ln(y)$. Hasil prediksi nantinya dikembalikan ke format Rupiah menggunakan fungsi eksponensial.

Pemodelan XGBoost

Model regresi dibangun menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) yang diintegrasikan ke dalam *pipeline* pemrosesan *scikit-learn*. Untuk mencapai performa optimal, parameter model diuji dan dikalibrasi menggunakan metode pencarian *Grid Search* yang dipadukan dengan skema *3-Fold Cross-Validation*. Eksplorasi ini difokuskan

pada ruang pencarian parameter utama algoritma, yaitu tingkat pembelajaran (*learning rate*), kedalaman pohon keputusan (*max depth*), dan jumlah estimator (*n_estimators*), guna mengidentifikasi kombinasi yang menghasilkan tingkat galat terendah.

Evaluasi Kinerja Mobil

Untuk menjamin objektivitas pengujian dan menghindarkan sistem dari bias kebetulan saat pemisahan data (*data splitting bias*), performa model dievaluasi secara ketat menggunakan protokol *10-Fold Cross-Validation* dengan pengacakan baris data secara dinamis (*shuffle=True*). Kemampuan generalisasi dan prediksi model diukur melalui tiga instrumen metrik statistik utama:

1. Mean Absolute Percentage Error (MAPE)

Digunakan untuk mengukur rata-rata persentase penyimpangan (kesalahan absolut) antara estimasi harga dari model dengan nilai harga pasar yang aktual. Metrik ini mempermudah interpretasi seberapa besar tingkat toleransi *error* dalam konteks bisnis.

2. Root Mean Squared Error (RMSE)

Digunakan untuk mengukur standar deviasi dari galat (selisih) prediksi. Melalui proses kuadrat, metrik ini memberikan penalti yang lebih besar terhadap prediksi yang melenceng jauh (*outlier*), dan hasilnya dikalkulasikan ke dalam satuan dasar (Rupiah).

3. Coefficient of Determination (R^2)

Digunakan untuk mengukur seberapa besar proporsi atau persentase variansi pada harga mobil yang dapat dijelaskan secara valid oleh fitur-fitur independen kendaraan di dalam model regresi.

Arsitektur Integrasi dan Logika Reverse Calculation

Perangkat lunak dibangun menggunakan arsitektur terpisah (*decoupled architecture*) untuk mendistribusikan beban komputasi. Antarmuka pengguna dan logika manajemen data dikembangkan menggunakan *framework* berbasis PHP (Laravel) dan MySQL. Komputasi kecerdasan buatan diisolasi ke dalam layanan mikro (*microservice*) berbasis Python. Kedua environment ini berkomunikasi secara seketika (*real-time*) melalui pertukaran data berformat JSON berbasis protokol komunikasi *REST API*.

Keluaran prediksi dari model AI yang berupa Harga Pasar Wajar (HP_{pred}) kemudian diproses lebih lanjut oleh *backend* Laravel menggunakan logika penetapan harga kalkulasi mundur (*reverse calculation*). Rekomendasi Harga Beli Maksimal (HB_{max}) diperoleh melalui formulasi:

$$HB_{max} = HP_{pred} - (M_{target} + B_{perbaikan})$$

Dimana M_{target} merepresentasikan target margin keuntungan dan $B_{perbaikan}$ adalah estimasi biaya restorasi unit. Setelah HB_{max} diperoleh, sistem akan mengevaluasi kelayakan bisnis dengan mengomparasinya terhadap Harga Penawaran (H_{tawar}) dari penjual. Jika $H_{tawar} \leq HB_{max}$, maka sistem akan mengeluarkan indikator kelayakan "AMAN". Arsitektur terintegrasi ini memungkinkan penaksir di lapangan menerima rentang estimasi wajar, batas harga tawar, dan rekomendasi keputusan secara komprehensif.

HASIL DAN PEMBAHASAN

Hasil Pra-pemrosesan Data

Tahapan awal penelitian ini memproses total populasi data mentah sebanyak 13.362 entri transaksi historis dan penawaran pasar. Penerapan algoritma pembersihan data dan identifikasi pencilan (*outlier*) menggunakan metode Grouped Interquartile Range (*Grouped-IQR*) berbasis segmen merek berhasil menyaring 1.038 baris data anomali. Metode spesifik domain ini terbukti efektif dalam membuang kesalahan entri atau harga spekulatif ekstrem tanpa merusak distribusi alami harga pada segmentasi mobil mewah. Rincian penyusutan jumlah dataset dapat dilihat pada Tabel 1.

Tabel 1. Ringkasan Pra-pemrosesan *Dataset*

Keterangan	Jumlah Data	Persentase
Data Mentah Awal (<i>Raw Data</i>)	13.362	100%
Data Terindikasi <i>Outlier</i>	1.038	7,76%
Data Bersih Siap Latih (<i>Clean Data</i>)	12.324	92,24%

Hasil akhir dari proses ini menghasilkan himpunan data bersih sebanyak 12.324 baris data yang siap digunakan untuk fase pelatihan dan pengujian model.

Optimasi Hyperparameter Model XGBoost

Untuk mencapai performa prediksi yang optimal, dilakukan *hyperparameter tuning* pada algoritma *Extreme Gradient Boosting* (XGBoost). Optimasi *hyperparameter* dilakukan dengan *Grid Search* yang dikombinasikan dengan validasi silang tiga lipatan (*3-Fold Cross-Validation*). Ruang pencarian (*search space*) dieksekusi terhadap 27 kombinasi *hyperparameter* yang melibatkan *learning_rate* (0.01, 0.05, 0.1), *max_depth* (3, 5, 7), dan *n_estimators* (300, 500, 1000). Hasil ekstraksi

komputasi komparatif dari kombinasi terbaik disajikan pada Tabel 2.

Tabel 2. Tiga Kombinasi *Hyperparameter* Terbaik

Peringkat	Learning Rate	Max Depth	N Estimators	MAPE (%)
1	0.1	5	1000	11.89%
2	0.1	5	500	12.09%
3	0.05	5	1000	12.12%

Berdasarkan Tabel 2, kombinasi *hyperparameter* terbaik dicapai pada konfigurasi: *learning_rate* = 0.1, *max_depth* = 5, dan *n_estimators* = 1000. Nilai *max_depth* yang moderat (5) memberikan keseimbangan arsitektur yang ideal; memungkinkan pohon keputusan untuk menangkap pola interaksi fitur yang cukup kompleks tanpa terjebak pada proses menghafal data latih (*overfitting*). Kompleksitas kedalaman pohon ini diimbangi dengan jumlah estimator yang tinggi (1000 iterasi) serta nilai *learning_rate* sebesar 0.1. Karakteristik ini memastikan proses minimasi galat (*residuals*) berlangsung secara iteratif dan stabil terhadap data fluktuasi harga pasar.

Evaluasi Kinerja Model

Sebelum mengevaluasi performa model XGBoost yang telah dioptimasi, penelitian ini terlebih dahulu menetapkan model *baseline* (dasar) menggunakan algoritma Regresi Linier Berganda pada dataset yang sama. Pengujian *10-Fold Cross-Validation* pada model *baseline* tersebut menghasilkan nilai MAPE sebesar 21,62%, RMSE sebesar Rp 64.187.783, dan skor R^2 sebesar 0.6303. Performa *baseline* ini memvalidasi hipotesis awal penelitian bahwa algoritma regresi konvensional kesulitan memetakan pola *non-linear* pada proses depresiasi harga aset otomotif.

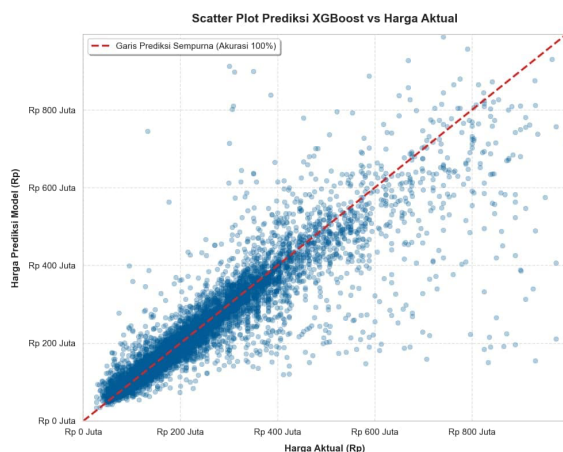
Sebagai perbandingan, keandalan dan stabilitas model XGBoost dengan parameter terbaik yang diusulkan selanjutnya divalidasi menggunakan skema *10-Fold Cross-Validation* yang sama. Pengujian ini mengacak baris data secara dinamis dan membaginya menjadi 10 lipatan sama besar untuk mengevaluasi konsistensi model. Ringkasan hasil evaluasi performa model disajikan pada Tabel 3.

Tabel 3. Hasil Pengujian *10-Fold Cross-Validation*

Lipatan Pengujian	MAPE (%)	RMSE (Rp)	Koef. Determinasi (R^2)
Fold 01	11.27%	Rp 62,708,977	0.7830
Fold 02	10.99%	Rp 57,038,702	0.8272

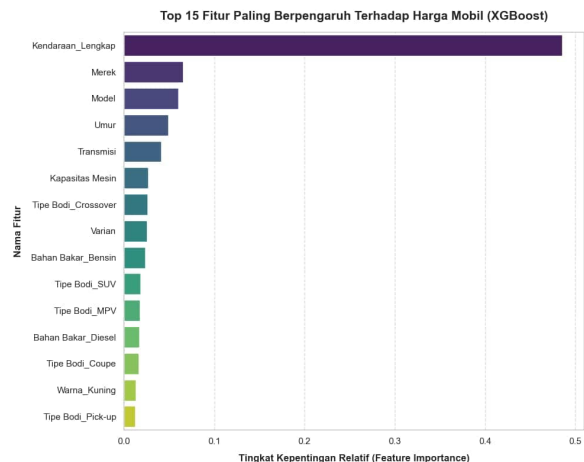
Fold 03	11.94%	Rp 63,129,153	0.8124
Fold 04	11.25%	Rp 53,624,452	0.8491
Fold 05	11.81%	Rp 62,998,255	0.7856
Fold 06	10.32%	Rp 46,209,265	0.8864
Fold 07	11.08%	Rp 50,357,769	0.8573
Fold 08	11.07%	Rp 49,534,958	0.8558
Fold 09	11.46%	Rp 47,983,882	0.8807
Fold 10	11.11%	Rp 54,208,958	0.8482
Rata-rata (Mean)	11.23%	Rp 54,779,437	0.8386
Standar Deviasi (±)	0.45%	Rp 6,442,836	0.0359

Secara statistik, Tabel 3 membuktikan ketangguhan algoritma XGBoost pada data pasar yang bervariasi. Nilai MAPE rata-rata mencapai 11,23%, mengindikasikan tingkat persentase kesalahan prediksi dari harga aktual sangat kecil dan terkendali. Sementara itu, nilai rata-rata RMSE sebesar Rp 54.779.437 menunjukkan standar deviasi galat prediksi dalam satuan mata uang aktual. Mengingat rentang spesifikasi dan harga objek penelitian (mobil bekas) yang merentang luas dari puluhan juta hingga skala miliaran rupiah, tingkat deviasi ini dikategorikan aman dan masih berada dalam batas toleransi operasional penetapan harga beli bisnis. Skor R² dengan rata-rata 0.8386 (83,86%) menegaskan kemampuan model yang tinggi dalam menjelaskan variansi data riil. Aspek krusial dari evaluasi ini terlihat pada nilai standar deviasi antar lipatan yang sangat minim ($\pm 0.45\%$ pada MAPE dan ± 0.0359 pada R²), menyimpulkan model tidak rentan terhadap bias penarikan sampel (*sampling bias*) dan konsisten dalam merumuskan prediksi.



Gambar 1. Perbandingan Hasil Prediksi Model dan Nilai Harga Aktual Pasar

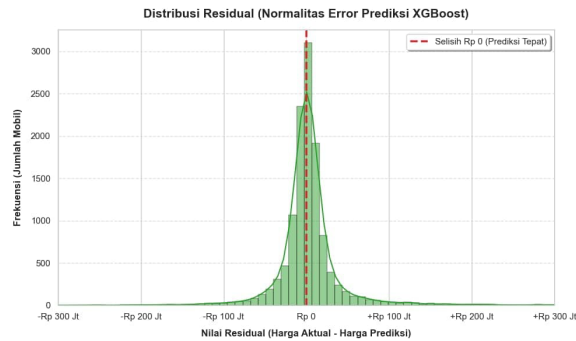
Visualisasi Scatter Plot pada Gambar 1 memperlihatkan kedekatan sebaran titik harga prediksi dengan garis putus-putus merah (Akurasi 100%). Kepadatan titik yang sangat tinggi di rentang harga Rp 0 hingga Rp 400 Juta membuktikan bahwa model sangat akurat dalam memprediksi nilai kendaraan pada segmen *entry-level* hingga menengah (*volume maker*). Penyebaran (*varians*) yang sedikit melebar pada rentang harga di atas Rp 600 Juta merupakan fenomena yang dapat dijelaskan secara empiris. Pada segmen mobil kelas atas (*premium*), rentang harga sangat dipengaruhi oleh faktor-faktor subjektif yang sulit direpresentasikan sepenuhnya oleh atribut tabular murni. Faktor tersebut meliputi penambahan kelengkapan aksesoris (*aftermarket*), histori perawatan (*service record*), hingga preferensi status kepemilikan. Ketiadaan variabel-variabel mikro tersebut di dalam dataset menyebabkan model mengalami sedikit pergeseran presisi mutlak pada segmen mewah, namun secara agregat, model tetap terbukti sangat andal untuk mendikte harga pasar secara umum.



Gambar 2. Tingkat Kepentingan Fitur (*Feature Importance*) terhadap Estimasi Harga

Lebih lanjut, analisis *Feature Importance* XGBoost (Gambar 2) mengungkapkan temuan empiris yang berharga terkait karakteristik pasar otomotif domestik. Secara klasik, "Umur" kendaraan seringkali dikategorikan sebagai faktor depresiasi utama. Namun, model justru menetapkan atribut rekayasa fitur "Kendaraan_Lengkap" (yang mengintegrasikan identitas spesifik Merek, Model, dan Varian) sebagai prediktor dengan bobot pengaruh paling dominan secara absolut. Hal ini secara presisi mencerminkan realita konsumen di mana varian atau *trim level* menjadi penentu utama terhadap fitur keselamatan, kenyamanan, dan retensi nilai aset (*hold value*). Sebagai contoh kasus, kendaraan keluaran tahun 2014 dengan tipe tertinggi (varian lengkap) terbukti mampu

mempertahankan nilai jual yang jauh lebih tinggi dibandingkan kendaraan yang sama keluaran tahun 2015 namun dengan tipe terendah. Fakta algoritmik ini memvalidasi bahwa penggabungan identitas Merek-Model-Varian menggunakan teknik Target Encoding merupakan tahapan pra-pemrosesan yang paling krusial dalam mendongkrak akurasi XGBoost.



Gambar 3. Distribusi Galat (Residual) Prediksi XGBoost

Analisis lanjutan terhadap distribusi galat (*residual distribution*) pada Gambar 3 menunjukkan bahwa selisih antara nilai prediksi dan aktual terdistribusi secara normal di sekitar nilai nol (0) membentuk kurva lonceng yang simetris. Hal ini mengonfirmasi bahwa model XGBoost yang dibangun tidak memiliki kecenderungan bias yang sistematis (tidak secara konsisten melakukan *over-predicting* maupun *under-predicting*), sehingga memenuhi asumsi dasar pemodelan regresi yang ideal.

Perbandingan Kinerja dengan Penelitian Terdahulu

Untuk mengukur signifikansi optimasi yang dilakukan, performa model yang dibangun dibandingkan dengan metode regresi lain dari berbagai studi terdahulu, seperti yang terlihat pada Tabel 4.

Tabel 4. Perbandingan Kinerja Prediksi dengan Penelitian Terdahulu

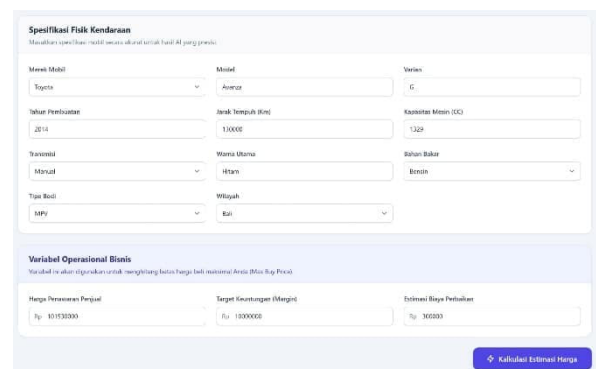
Peneliti (Tahun)	Algoritma Model	Nilai Evaluasi (R ²)
Ghaisani & Baita (2025)	K-Nearest Neighbor (KNN)	Bias pada sebaran varians tinggi
Idris & Pontooyo (2025)	Regresi Linear Berganda	0.4179
F. Sarasati, dkk. (2025)	XGBoost (Dataset Properti)	0.7700
Penelitian Ini (2026)	XGBoost (Terintegrasi Logika Web)	0.8423

Pemilihan koefisien determinasi (R²) sebagai metrik komparasi utama pada Tabel 4 didasarkan pada sifatnya yang *scale-independent* (tidak terpengaruh oleh skala besaran data). Berbeda dengan metrik galat nominal seperti RMSE yang mutlak bergantung pada rentang harga *dataset* masing-masing domain (misalnya selisih harga properti dalam miliaran berbanding mobil dalam puluhan juta), evaluasi R² memberikan tolok ukur yang universal dan objektif untuk membandingkan kecocokan variansi antar model regresi.

Komparasi performa ini membuktikan secara konklusif validitas dan kebaruan (*novelty*) dari pendekatan yang diusulkan. Angka R² sebesar 0.8386 yang dicapai pada penelitian ini bukanlah sekadar hasil langsung dari penggunaan algoritma XGBoost standar, melainkan merupakan sintesis keberhasilan dari arsitektur pra-pemrosesan yang didesain secara spesifik (*domain-aware*). Penerapan *Grouped-IQR* terbukti krusial dalam menyelamatkan data transaksi mobil segmen premium dari penghapusan anomali yang keliru akibat filter statistik global. Di samping itu, teknik Target Encoding berhasil mengekstrak esensi nilai ekonomis dari ribuan kombinasi varian kendaraan secara efisien tanpa memicu masalah ledakan dimensi (*curse of dimensionality*). Sinergi metode inilah yang menjadikan model prediksi harga yang dibangun memiliki tingkat adaptabilitas yang jauh lebih unggul dibandingkan dengan pendekatan regresi linier pada literatur terdahulu.

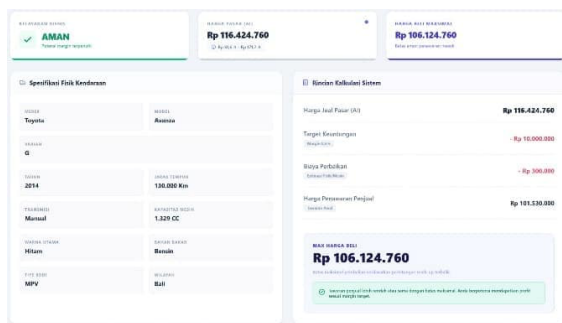
Implementasi Sistem dan Logika Reverse Calculation

Nilai lebih penelitian ini dibandingkan literatur sebelumnya adalah arsitektur terapan (*applied implementation*). Model kecerdasan buatan di-*deploy* sebagai layanan mikro (API Python) yang terintegrasi secara real-time ke dalam sistem antarmuka berbasis web (Laravel) CV Guna Abadi.



Gambar 4. Antarmuka Penginputan Spesifikasi Fisik dan Variabel Bisnis

Seperti yang ditunjukkan pada Gambar 4, penaksir menginputkan spesifikasi kendaraan (contoh: Toyota Avanza G 2014, 130.000 Km) beserta variabel operasional bisnis, yaitu harga tawaran penjual (Rp 101.530.000), target margin keuntungan (Rp 10.000.000), dan estimasi biaya perbaikan (Rp 300.000). Laravel kemudian mentransmisikan data fisik kendaraan ke layanan *Machine Learning*.



Gambar 5. Dashboard Rincian Kalkulasi dan Kelayakan Bisnis

Setelah model mengembalikan nilai Harga Pasar Wajar sebesar Rp 116.424.760, sistem web mengeksekusi logika kalkulasi mundur (*Reverse Calculation*). Berdasarkan Gambar 5, sistem mereduksi harga pasar wajar dengan target margin dan biaya perbaikan, menghasilkan formulasi *Max* Harga Beli sebesar Rp 106.124.760. Lebih lanjut, sistem memvalidasi kelayakan secara otomatis. Karena Harga Penawaran Penjual (Rp 101.530.000) berada di bawah Batas Harga Beli Maksimal, Dasbor secara cerdas menampilkan indikator Kelayakan Bisnis "AMAN".

Arsitektur aplikasi terpadu ini menuntaskan problem lambatnya penentuan batas tawar yang dialami subjek riset. Keandalan arsitektur ini juga telah divalidasi melalui pengujian performa sistem (*performance testing*) secara *end-to-end*. Berdasarkan hasil pengujian latensi, model XGBoost ter-serialisasi yang berukuran sangat ringan (0,72 MB) mampu melakukan inferensi murni dalam waktu rata-rata 14,62 milidetik. Adapun total waktu respons arsitektur, yang mencakup pengiriman *request HTTP* dari *frontend* Laravel, pemrosesan di layanan mikro API Python, hingga eksekusi kalkulasi mundur (*reverse calculation*) untuk menghasilkan rekomendasi akhir, tercatat hanya memakan waktu rata-rata 2,07 detik. Waktu komputasi yang sangat efisien ini secara empiris membuktikan bahwa sistem beroperasi tanpa hambatan dan sepenuhnya layak digunakan sebagai instrumen taktis *real-time* (*user-friendly*) oleh penaksir aset di lapangan guna mempersempit area bias kognitif dan memitigasi risiko kerugian sejak dini.

KESIMPULAN

Berdasarkan pencapaian metrik empiris dengan nilai rata-rata *Mean Absolute Percentage Error* (MAPE) sebesar 11,23%, *Root Mean Squared Error* (RMSE) sebesar Rp 54.779.437, dan skor Koefisien Determinasi (R^2) sebesar 0.8386, dapat disimpulkan bahwa algoritma *Extreme Gradient Boosting* (XGBoost) terbukti sangat tangguh dan efektif dalam memprediksi harga pasar wajar mobil bekas secara objektif. Capaian akurasi prediksi ini tidak terlepas dari penerapan inovasi pra-pemrosesan data berbasis domain. Penggunaan metode *Grouped-IQR* dan Target Encoding terbukti krusial dalam mempertahankan integritas sebaran harga kendaraan premium sekaligus menangani variabel berkardinalitas tinggi secara efisien, yang pada akhirnya membawa performa XGBoost jauh melampaui algoritma regresi konvensional pada studi terdahulu. Lebih lanjut, fungsionalitas algoritma AI yang rumit tersebut berhasil diintegrasikan ke dalam arsitektur perangkat lunak berbasis web melalui layanan mikro API Python dengan waktu latensi *end-to-end* yang sangat responsif (2,07 detik). Arsitektur ini memungkinkan sistem untuk mentransformasi prediksi harga pasar menjadi rekomendasi batas Harga Beli Maksimal secara *real-time* melalui mekanisme *reverse calculation*, yang secara praktis memberikan solusi terstandarisasi untuk meminimalisir risiko kerugian finansial akibat bias taksiran manual pada entitas bisnis.

DAFTAR PUSTAKA

- Ariyono, L. D., & Manafe, L. A. (2022). Penetapan Harga Jual Produk Terhadap Volume Penjualan Pada CV Vulkanindo Mitra Abadi. *Business and Entrepreneurship Journal (BEJ)*, 3(1), 40–47.
- Arradian, D. (2025, January 22). *Gaikindo: Penjualan Mobil Baru 865.723 Unit, Mobil Bekas 1,8 Juta di 2024*. SindoNews. <https://otomotif.sindonews.com/>
- Bolikulov, F., Nasimov, R., Rashidov, A., Akhmedov, F., & Cho, Y. I. (2024). Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms. *Mathematics*, 12(16). <https://doi.org/10.3390/math12162553>
- Eka Adinagoro, R., Dwi Rangga Premana, I., Bintang Pamungkas, R., Aryasetya, P., & Joedhistiro, D. (2025). Pengembangan Sistem Prediksi Harga Mobil Bekas Di Pasar India Menggunakan Algoritma XGBoost dan NLP. *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis (SENATIB) 2025*, 207–215. <https://doi.org/10.47701/w6511j19>
- Fayyaz, I., Ali, G. G. M. N., & Khairunnesa, S. S. (2025). Advanced Feature Engineering and Machine Learning Techniques for High Accurate Price Prediction of Heterogeneous Pre-

- Own Cars. *Vehicles*, 7(3), 1–30.
<https://doi.org/10.3390/vehicles7030094>
- Ghaisani, M. S., & Baita, A. (2025). Implementation of the K-Nearest Neighbors (KNN) Regressor Method to Predict Toyota Used Car Prices. *Journal of Applied Informatics and Computing (JAIC)*, 9(1), 194–201.
- Haya, F., & Nudin, S. R. (2024). Sistem Informasi Prediksi Harga Mobil Bekas Menggunakan Metode KNN. *Jurnal Manajemen Informatika (JMI)*, 13(02), 8–12.
- Idris, N. O., & Pontooyo, F. (2025). Evaluasi Model Machine Learning untuk Prediksi Harga Mobil dengan Perbandingan Ensemble dan Regresi Linear. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIRSI)*, 4(1), 129–143.
<https://doi.org/10.70340/jirsi.v4i3>
- Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Computational Statistics*, 37(5), 2671–2692.
<https://doi.org/10.1007/s00180-022-01207-6>
- Safitri, O., Mikhratunnisa, & Rizqi, R. M. (2024). Model Gompertz pada Depresiasi Aset Kendaraan Mobil Honda Brio. *Jurnal Ilmiah Raflesia Akuntansi*, 10(2), 988–994.
<https://doi.org/10.53494/jira.v10i2.680>
- Syafarina, G. A., Amin, M., & Zaenuddin. (2026). Perbandingan Tree-Based Ensemble (Random Forest, XGBoost, LightGBM) dan Deep Learning untuk Prediksi Klik Iklan Online. *Jurnal Technologia (JIT)*, 17(1), 257–265.
<https://doi.org/10.31602/tji.v17i1.21627>
- Winata, A. J. (2024). Prediksi Harga Mobil Bekas Menggunakan Algoritma Gradient Boosting Machine dan Random Forest. *Jurnal Inovasi Pendidikan Kreatif*, 5(4), 454–464.
- Yang, J., Kim, J., Ryu, H., Lee, J., & Park, C. (2024). Predicting Car Rental Prices: A Comparative Analysis of Machine Learning Models. *Electronics (Switzerland)*, 13(12), 1–20.
<https://doi.org/10.3390/electronics13122345>