

Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
<https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index>
 Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
 Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
 e-ISSN: 2685-0893

Implementasi Sistem Prediksi Harga Motor Bekas Menggunakan Algoritma *Categorical Boosting*

DOI: <http://dx.doi.org/10.35889/jutisi.v15i3.3730>

Creative Commons License 4.0 (CC BY – NC)



**Fathia Wardah S. Djawas^{1*}, Muhammad Najamuddin Dwi Miharja²,
 Nanang Tedi K³**

Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

*e-mail *Corresponding Author*: ilhamwahyudi4625@gmail.com

Abstract

This study is intended to design a web-based system for predicting used motorcycle prices by employing the Categorical Boosting (CatBoost) algorithm. The research problem stems from the subjective process of setting used motorcycle prices, which is affected by multiple factors and frequently results in inaccurate price estimates. The research methodology covers data collection from the OLX Indonesia platform, followed by data preprocessing, feature engineering, data partitioning with a train–test split approach, and model construction using the CatBoost Regressor algorithm. The experimental results indicate that the prediction model attains a coefficient of determination of 0.9166 on the training set and 0.9149 on the test set. These findings suggest that the model performs well, accounting for more than 91% of the variance in used motorcycle prices and producing stable price predictions without notable overfitting. The system is deployed using Streamlit, enabling users to obtain used motorcycle price predictions in a faster, interactive, and more objective manner.

Keywords: *Used Motorcycle; Machine Learning; CatBoost; Streamlit*

Abstrak

Penelitian ini dimaksudkan untuk merancang sebuah sistem prediksi harga motor bekas berbasis web dengan memanfaatkan algoritma *Categorical Boosting* (CatBoost). Permasalahan penelitian berangkat dari praktik penentuan harga motor bekas yang masih bersifat subjektif dan dipengaruhi beragam faktor sehingga kerap menghasilkan estimasi harga yang kurang akurat. Metode yang digunakan mencakup pengumpulan data dari platform OLX Indonesia, *preprocessing data*, *feature engineering*, pembagian data menggunakan skema *train–test split*, serta pemodelan dengan algoritma *CatBoost Regressor*. Hasil pengujian memperlihatkan bahwa model prediksi memperoleh nilai koefisien determinasi (R^2) sebesar 0,9166 pada data latih dan 0,9149 pada data uji. Temuan ini mengindikasikan bahwa model memiliki kinerja yang baik, karena mampu menjelaskan lebih dari 91% variasi harga motor bekas serta menghasilkan prediksi yang konsisten tanpa gejala *overfitting* yang berarti. Sistem kemudian diimplementasikan menggunakan Streamlit sehingga pengguna dapat memperoleh estimasi harga motor bekas dengan lebih cepat, interaktif, dan bersifat lebih objektif.

Kata kunci: *Prediksi Harga; Motor Bekas; Machine Learning; CatBoost; Streamlit*

1. Pendahuluan

Perkembangan teknologi informasi serta pembelajaran mesin dalam beberapa tahun terakhir telah mendorong pemanfaatan pendekatan berbasis data di berbagai sektor, termasuk pada industri otomotif. Sepeda motor merupakan salah satu jenis transportasi utama bagi masyarakat Indonesia karena dianggap lebih ekonomis, efisien, dan fleksibel dalam menunjang aktivitas sehari-hari [1]. Tingginya jumlah dan distribusi sepeda motor di Indonesia berdampak pada meningkatnya aktivitas jual beli kendaraan bekas, khususnya di wilayah perkotaan dan daerah penyangga seperti Bekasi. Kondisi ini mendorong pentingnya sistem yang mampu

membantu menentukan harga motor bekas secara akurat dan objektif guna mendukung transaksi yang adil dan transparan [2].

Dalam praktiknya, penentuan harga motor bekas masih banyak dilakukan secara subjektif berdasarkan pengalaman penjual atau kondisi pasar yang fluktuatif. Harga kendaraan ditentukan oleh sejumlah faktor seperti merek, model kendaraan, tahun produksi, kapasitas mesin, kondisi fisik, serta jarak tempuh, sehingga proses estimasi harga tanpa dukungan model prediktif cenderung kurang konsisten [3]. Penelitian lain juga mengindikasikan bahwa regresi linear mampu dimanfaatkan guna mengestimasi nilai jual kendaraan bekas dengan memanfaatkan kombinasi berbagai atribut kendaraan, meskipun hasilnya masih bergantung pada asumsi hubungan linier antarvariabel [4]. Kompleksitas atribut dan dinamika pasar menyebabkan masyarakat sering mengalami kesulitan dalam menerapkan harga jual maupun beli kendaraan yang mencerminkan kondisi sebenarnya [5].

Berbagai penelitian sebelumnya telah mengkaji prediksi harga kendaraan bekas menggunakan beragam metode. Pendekatan regresi linear dilaporkan mampu memberikan hasil yang cukup baik dalam estimasi harga kendaraan, baik pada konteks mobil maupun motor bekas [6]. Namun, beberapa studi menunjukkan bahwa metode ini memiliki keterbatasan dalam menggambarkan hubungan nonlinier dan interaksi lebih kompleks [7]. Di sisi lain, penelitian lain memanfaatkan metode *time series* seperti *Winter Exponential Smoothing* untuk meramalkan penjualan sepeda motor bekas, sehingga fokusnya lebih kepada jumlah penjualan daripada estimasi harga kendaraan [8]. Metode klasifikasi seperti *Naive Bayes* juga telah diterapkan untuk memprediksi keputusan pembelian motor bekas, sehingga keluaran yang dihasilkan berupa kategori keputusan, bukan nilai harga numerik [9]. Selain itu, pendekatan berbasis *Artificial Neural Network* dilaporkan mampu meningkatkan akurasi prediksi harga kendaraan bekas [10]. Penelitian lain berbasis machine learning juga menunjukkan bahwa penggunaan berbagai model dan dataset yang lebih kompleks dapat meningkatkan performa prediksi secara signifikan [11]. Lebih lanjut, studi lain menekankan pentingnya pemilihan fitur yang tepat dalam meningkatkan akurasi model prediksi harga kendaraan [12].

Berdasarkan perkembangan tersebut, penelitian ini mengusulkan pembangunan sistem prediksi harga motor bekas berbasis web menggunakan algoritma *Categorical Boosting (CatBoost)*. Algoritma ini merupakan bagian dari keluarga *gradient boosting* dan dirancang untuk menangani data kategorikal secara lebih efektif, sehingga sesuai untuk data tabular kendaraan yang terdiri atas kombinasi atribut numerik dan kategorikal. Dalam beberapa penelitian sebelumnya, model *regresi linear* masih banyak digunakan karena kemudahan dalam interpretasi serta implementasinya yang relatif sederhana [13]. Namun, perkembangan metode machine learning seperti *XGBoost* dan *LSTM* mengindikasikan bahwa algoritma berbasis *boosting* lebih efektif dalam menangani hubungan data yang memiliki pola kompleks dan nonlinier [14]. Selain itu, beberapa studi sebelumnya juga mengindikasikan bahwa integrasi model *machine learning* ke dalam aplikasi prediksi berbasis web dapat memberikan kemudahan bagi pengguna dalam memperoleh estimasi harga kendaraan secara cepat dan interaktif [15]. Dengan memanfaatkan atribut kendaraan seperti merek, model, tahun produksi, kapasitas mesin, jarak tempuh, dan kondisi kendaraan, model *CatBoost Regressor* dibangun dan diintegrasikan ke dalam sistem berbasis web menggunakan *Streamlit*. Kebaruan penelitian ini berfokus pada penerapan *algoritma CatBoost* untuk memperkirakan harga motor bekas di Indonesia, sehingga diharapkan dapat menghasilkan perkiraan harga yang lebih akurat, konsisten, dan mudah dimanfaatkan oleh masyarakat.

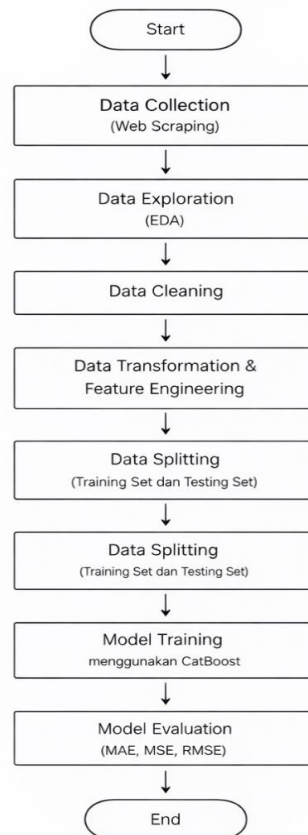
2. Metodologi

Penelitian ini menggunakan pendekatan pembelajaran mesin untuk mengembangkan sistem perkiraan harga kendaraan bekas berbasis web menggunakan algoritma *Categorical Boosting (CatBoost)*. Tahapan penelitian dilaksanakan secara sistematis mulai dari pengumpulan data, pengolahan dan analisis, hingga penarikan Kesimpulan tahap pembangunan, pengujian, serta evaluasi sistem prediksi harga sepeda motor bekas berbasis web.

2.1 Tahapan Penelitian

Tahapan penelitian dilakukan melalui beberapa proses, yaitu pengumpulan data, *preprocessing* data, *feature engineering*, pembagian data *training* dan *testing*, pemodelan

menggunakan algoritma *CatBoost*, evaluasi model, serta implementasi sistem berbasis web menggunakan *Streamlit*.



Gambar 1. Flowchart Tahapan Pengolahan Data dan Pemodelan

Flowchart pada Gambar 1 menunjukkan tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data, prapemrosesan, pembangunan model prediksi, hingga evaluasi dan penarikan kesimpulan, praproses data, pemodelan, hingga evaluasi kinerja, yang dilakukan dalam pembangunan sistem untuk memprediksi harga sepeda motor bekas dengan memanfaatkan algoritma *CatBoost*. Secara umum, alur penelitian meliputi pengumpulan data dilakukan melalui teknik web scraping, kemudian dilanjutkan dengan tahap *analisis data eksploratif (EDA)*, *pembersihan dataset*, *transformasi serta feature engineering*, *pemisahan data training dan testing*, *proses training model*, serta *evaluasi kinerja model dengan MAE, MSE, dan RMSE*.

2.2 Pengumpulan Data

Dataset penelitian didapatkan dari platform OLX Indonesia melalui teknik pengambilan data *web scraping*. Dataset yang digunakan berisi informasi kendaraan bekas seperti merek kendaraan, model kendaraan, tahun produksi, kapasitas mesin, jarak tempuh, jenis penjual, dan harga kendaraan. Data yang diperoleh kemudian diseleksi dan disesuaikan dengan kebutuhan penelitian agar dapat digunakan pada proses pemodelan.

2.3 Tahap Prapemrosesan Data

Tahap prapemrosesan data dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam pelatihan model. Proses ini meliputi penanganan data yang hilang (*missing value*), penghapusan data yang berulang, penanganan outlier, data proses transformasi data berbentuk angka dan kategorikal agar dapat diproses secara optimal oleh *algoritma CatBoost*.

Pada tahap ini juga dilakukan proses *feature engineering* untuk menghasilkan fitur baru yang dapat meningkatkan performa model prediksi. Salah satu proses yang dilakukan adalah transformasi tahun kendaraan menjadi umur kendaraan serta penyesuaian format atribut numerik seperti jarak tempuh dan kapasitas mesin.

Tabel 1. Fitur yang Digunakan dalam Pemodelan

No	Fitur	Keterangan
1	Merek	Merek Kendaraan
2	Model	Tipe kendaraan
3	Tahun Kendaraan	Tahun produksi motor
4	Kapasitas Mesin	Kapasitas mesin kendaraan
5	Jarak Tempuh	Total kilometer kendaraan
6	Jenis Penjual	Tipe penjual kendaraan
7	Harga	Target Prediksi

2.4 Pembagian Data

Dataset yang telah melalui tahap prapemrosesan kemudian dipisahkan ke dalam data pelatihan (training) dan data pengujian (testing) dengan menerapkan metode train–test split. Proses pembagian data dilakukan dengan menggunakan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Data pelatihan digunakan pada tahap pembentukan model, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model dalam melakukan prediksi terhadap data baru.

Tabel 2. Distribusi Data *Training* dan *Testing*

Jenis Data	Persentase
Training	80%
Testing	20%

2.5 Implementasi Algoritma *CatBoost*

Algoritma *Categorical Boosting (CatBoost)* digunakan sebagai metode utama prediksi harga motor bekas karena mampu mengolah data kategorikal tanpa *encoding* yang kompleks dan memanfaatkan *gradient boosting* untuk mengurangi galat prediksi. Secara umum, pembaruan model *CatBoost* pada iterasi ke- m dapat dituliskan sebagai berikut.

$$F_{m(x)} = F_{\{m-1\}(x)} + \alpha \cdot h_{m(x)} \quad (1)$$

Keterangan:

$F_m(x)$: Model prediksi pada iterasi ke- m

$F_{\{m-1\}(x)}$: Model prediksi sebelumnya

α : *Learning rate*

$h_m(x)$: *Weak learner* berupa pohon keputusan

x : Data input atau fitur

Tabel 3. Parameter Model *CatBoost*

Parameter	Nilai
Learning Rate	0,1
Iteration	1000
Depth	6
Loss Function	RMSE

2.6 Pengujian Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kinerja algoritma *CatBoost* dalam memprediksi harga motor bekas. Pengujian dilakukan dengan menggunakan beberapa metrik evaluasi regresi, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, serta koefisien determinasi (R^2).

Persamaan *Mean Absolute Error (MAE)* ditunjukkan sebagai berikut:

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

Persamaan *Mean Square Error (MSE)* ditunjukkan sebagai berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

Persamaan *Root Mean Squared Error (RMSE)* ditunjukkan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

Semakin kecil nilai *MAE*, *MSE*, dan *RMSE*, maka performa model prediksi semakin baik. Selain itu, nilai koefisien determinasi (R^2) berfungsi untuk mengukur seberapa baik model mampu menjelaskan variasi yang terdapat pada data. pada variable terikat yang di sebabkan oleh variable bebas.

2.7 Implementasi Sistem

Model prediksi yang telah dikembangkan kemudian diterapkan ke dalam sistem berbasis web dengan menggunakan framework Streamlit. Sistem ini memungkinkan pengguna untuk memasukkan atribut kendaraan seperti merek, model, tahun kendaraan, kapasitas mesin, serta jarak tempuh guna memperoleh hasil prediksi harga motor bekas secara cepat dan interaktif.

3. Persiapan dan Prapemrosesan Data

3.1 Prapemrosesan Data

Proses *preprocessing* dilakukan untuk meningkatkan kualitas dataset sebelum digunakan pada tahap pelatihan model. Tahapan ini mencakup kegiatan pembersihan data, transformasi pada atribut numerik, serta pemilihan fitur yang akan diterapkan dalam proses pemodelan. Tujuan dari tahapan ini adalah agar data menjadi lebih konsisten sehingga dapat meningkatkan kinerja model dalam memprediksi harga motor bekas.

Dataset penelitian ini memuat beberapa atribut kendaraan, antara lain merek, model, tahun produksi, kapasitas mesin (cc), jarak tempuh, tipe penjual, serta kelengkapan dokumen kendaraan. Contoh sampel dataset yang digunakan dalam penelitian ini ditampilkan pada 4.

Tabel 4. Contoh Sampel Data Motor Bekas

No	Merek	Model	Tahun	Jarak Tempuh (km)	Harga (Rp)
1	Honda	Vario 125	2019	25.000	17.500.000
2	Honda	Beat	2020	35.000	12.000.000
3	Honda	PCX 150	2021	18.000	30.000.000
4	Yamaha	NMAX 155	2022	18.500	29.000.000
5	Yamaha	Aerox	2019	22.000	24.500.000

Data pada Tabel 4 menunjukkan bahwa atribut utama seperti merek, model, tahun kendaraan, jarak tempuh, dan harga telah tersedia dalam format yang siap digunakan pada tahap *preprocessing* dan pemodelan menggunakan algoritma *CatBoost*.

3.2 Implementasi Algoritma *CatBoost*

Pada bagian ini disajikan contoh hasil prediksi harga motor bekas yang dihasilkan oleh model *CatBoost* pada beberapa sampel data uji. Berbeda dengan Tabel 4 yang menampilkan contoh data awal kendaraan, pada bagian ini fokus pembahasan diarahkan pada hasil prediksi model dengan menyajikan identitas kendaraan berupa merek, model, dan tahun, serta evaluasi model dilakukan dengan membandingkan nilai harga aktual dengan hasil prediksi pada data pengujian

Nilai harga prediksi diperoleh setelah data kendaraan pada data *testing* diproses oleh model *CatBoost* yang telah dilatih pada tahap sebelumnya. Harga aktual merupakan nilai harga kendaraan yang tercatat pada data uji, sedangkan harga prediksi merupakan estimasi yang dihasilkan model berdasarkan pola hubungan antara atribut kendaraan dan harga yang dipelajari selama proses pelatihan. Selisih antara harga aktual dan harga prediksi digunakan untuk melihat tingkat kedekatan hasil prediksi model terhadap skor sebenarnya, yang semakin kecil selisihnya semakin baik kemampuan estimasinya.

Tabel 5. Contoh Hasil Prediksi Harga Motor Bekas

No	Merek	Model	Tahun	Harga Aktual (Rp)
1	Honda	Vario 125	2019	10.500.000
2	Honda	Beat	2020	19.800.000
3	Honda	PCX 150	2021	20.500.000
4	Yamaha	NMAX 155	2022	23.500.000
5	Yamaha	Aerox	2019	17.600.000

Tabel 5 menyajikan beberapa sampel kendaraan, seperti Honda Vario 125 tahun 2019, Honda Beat tahun 2020, Honda PCX 150 tahun 2021, Yamaha NMAX 155 tahun 2022, dan Yamaha Aerox tahun 2019, beserta harga aktual, harga prediksi, dan selisihnya. Beberapa sampel menunjukkan selisih yang relatif kecil, misalnya selisih sebesar Rp233.402 dan Rp268.338, yang mengindikasikan bahwa model mampu memberikan estimasi harga yang cukup dekat dengan nilai sebenarnya.

Secara umum, hasil prediksi pada beberapa sampel tersebut menunjukkan bahwa algoritma *CatBoost* mampu menangkap pola hubungan antara fitur kendaraan dan harga jual. Hal ini selaras dengan hasil studi-studi sebelumnya evaluasi model data *testing* yang memperlihatkan nilai R^2 sebesar 0,9149, *MAE* sebesar Rp679.370, dan *RMSE* sebesar Rp919.323, sehingga model dinilai memiliki performa prediksi yang baik dan stabil

3.3 Hasil Evaluasi Model

Evaluasi model dilakukan untuk menilai kemampuan model dalam memprediksi harga motor bekas pada data uji.

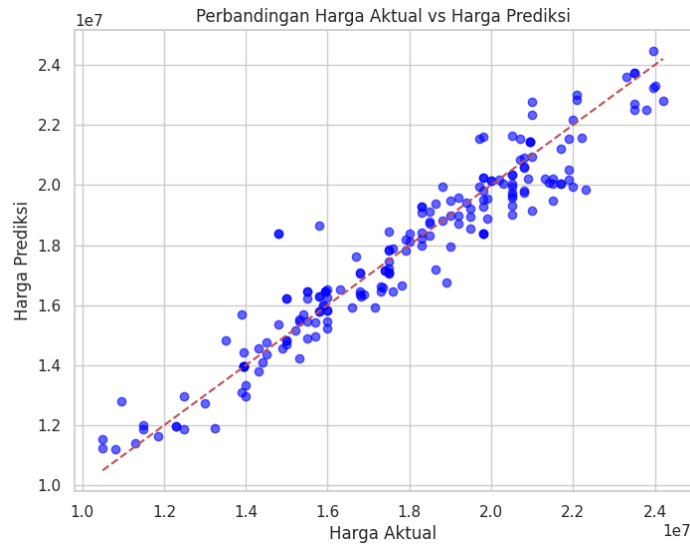
Nilai-nilai pada Tabel 6 diperoleh dengan melakukan perbandingan terhadap harga yang sebenarnya serta nilai harga hasil prediksi yang dihsikan oleh model *CatBoost*. Pada data *training*, model dimanfaatkan untuk menghasilkan perkiraan pada data latih, kemudian hasil estimasi tersebut kemudian dibandingkan dengan harga yang sebenarnya aktual untuk menghitung nilai *MAE*, *MSE*, *RMSE*, dan R^2 . Prosedur yang sama juga dilakukan pada data *testing*, yaitu dengan membandingkan harga aktual pada data uji dengan hasil prediksi model pada data yang tidak termasuk dalam proses pelatihan sebelumnya.

Tabel 6. Hasil Evaluasi Model *CatBoost*

Dataset	MAE	MSE	RMSE	R^2
<i>Training</i>	655.410	841.960.685.361	917.584	0,9166
<i>Testing</i>	679.370	845.155.266.364	919.323	0,9149

Berdasarkan Tabel 6, nilai *MAE* pada data *training* sebesar 655.410 dan pada data *testing* sebesar 679.370 menunjukkan bahwa rata-rata selisih absolut antara harga aktual dan harga prediksi berada pada rentang yang relatif serupa. Nilai *RMSE* pada data *training* sebesar 917.584 dan pada data *testing* sebesar 919.323 juga menunjukkan bahwa tingkat kesalahan prediksi model cenderung konsisten pada kedua kelompok data tersebut.

Nilai koefisien determinasi (R^2) sebesar 0,9166 pada data latih dan 0,9149 pada data uji menunjukkan bahwa model mampu menjelaskan lebih dari 91% variasi harga motor bekas berdasarkan variabel yang digunakan dalam studi ini.



Gambar 2. Perbandingan nilai harga actual dan estimasi model pada data pengujian.

Gambar 2 menampilkan hubungan antara nilai harga rill dan estimasi model pada data *testing* dalam bentuk *scatter plot*. Pada grafik tersebut, sumbu (x) menggambarkan harga rill, sedangkan sumbu (y) menggambarkan besaran harga prediksi yang merupakan keluaran dari model *CatBoost*.

Setiap titik pada Gambar 2 merepresentasikan satu sampel data kendaraan pada data *testing*. Semakin dekat posisi titik-titik yang berada pada garis diagonal, maka semakin kecil perbedaan antara harga nilai sebenarnya dan harga perkiraan, sedangkan titik yang berada cukup jauh dari garis acuan diagonal menggambarkan adanya kesalahan hasil estimasi yang lebih tinggi. Sebaran titik yang cenderung mengikuti garis diagonal menunjukkan bahwa sebagian besar hasil estimasi model mendekati nilai rill, sejalan dengan nilai R^2 pada data *testing* sebesar 0,9149, sehingga model *CatBoost* dinilai memiliki performa yang baik sebagai alat bantu estimasi harga motor bekas.

3.4 Implementasi Sistem

Model prediksi yang telah dibangun lalu diterapkan dalam sebuah system berbasis web dengan menggunakan framework *Streamlit*. Sistem memberikan kemampuan kepada pengguna melakukan prediksi harga motor bekas secara cepat dan interaktif berdasarkan atribut kendaraan yang dimasukkan.

The screenshot shows the "MotorPred" web application interface. At the top, there is a blue header with the logo and the text "Sistem Prediksi Harga Motor Bekas Menggunakan Algoritma CatBoost". Below the header, there are two tabs: "Prediksi Harga" (selected) and "About". The main content area is titled "Input Data Kendaraan" and contains a form with the following fields:

Kendaraan	Spesifikasi	Dokumen
Merek Motor Honda	Kapasitas Mesin (CC) 125	STNK Ada
Model Motor Beat	Jarak Tempuh (KM) 12000	BPKB Ada
Transmisi Automatic	Tipe Penjual Individu	Faktur Ada
Tahun Produksi 2023	Status Kepemilikan Ke-1	Buku Servis Ada

At the bottom of the form, there is a button labeled "Prediksi Harga Motor".

Gambar 3. Halaman Input Prediksi Harga Motor Bekas

Gambar 3 menunjukkan tampilan halaman input pada sistem prediksi harga motor bekas berbasis *web*. Pengguna dapat memasukkan informasi kendaraan seperti merek kendaraan, model kendaraan, tahun produksi, kapasitas mesin, dan jarak tempuh sebelum proses prediksi dilakukan oleh sistem.

The screenshot displays the 'MotorPred' web application interface. At the top, there is a blue header with the logo and the text 'Sistem Prediksi Harga Motor Bekas Menggunakan Algoritma CatBoost'. Below the header, there are two radio buttons: 'Prediksi Harga' (selected) and 'About'. The main section is titled 'Input Data Kendaraan' and contains a form with three columns: 'Kendaraan', 'Spesifikasi', and 'Dokumen'. The form fields are as follows:

Kendaraan	Spesifikasi	Dokumen
Merek Motor: Honda	Kapasitas Mesin (CC): 125	STNK: Ada
Model Motor: Beat	Jarak Tempuh (KM): 12000	BPKB: Ada
Transmisi: Automatic	Tipe Penjual: Individu	Faktur: Ada
Tahun Produksi: 2023	Status Kepemilikan: Ke-1	Buku Servis: Ada

Below the form, there is a button labeled 'Prediksi Harga Motor'. The result section, titled 'Estimasi Harga Motor Bekas', shows a large blue box with the estimated price: **Rp 14.271.121**. Below this, a price range is shown: Rentang: Rp 13.557.565 – Rp 14.984.678. At the bottom, there are three summary cards: 'Umur Motor 3 Tahun', 'Kapasitas Mesin 125 CC', and 'Jarak Tempuh 12.000 KM'. The footer text reads: 'MotorPred © 2026 | Implementasi Sistem Prediksi Harga Motor Bekas'.

Gambar 4. Halaman Hasil Prediksi Harga Motor Bekas

Gambar 4 menunjukkan hasil perkiraan harga motor bekas yang dihasilkan oleh sistem menggunakan algoritma *CatBoost*. Sistem menampilkan estimasi harga kendaraan berdasarkan data input pengguna secara cepat dan interaktif.

3.5 Pembahasan

Hasil evaluasi mengindikasikan bahwa model menghasilkan nilai koefisien determinasi (R^2) yang mencapai 0,9166 pada data latih dan 0,9149 pada data uji. Nilai tersebut yang menandakan model memiliki kemampuan yang berkinerja dalam memprediksi harga motor bekas serta tidak mengalami *overfitting* yang signifikan. Implementasi platform berbasis web juga dapat menghasilkan perkiraan harga kendaraan dengan cepat dan bersifat interaktif sehingga membantu pengguna dalam mendapatkan informasi harga yang lebih objektif serta akurat.

Temuan penelitian ini sejalan dengan penelitian terdahulu yang menyatakan bahwa pendekatan *ensemble learning* menunjukkan kinerja prediksi yang baik terhadap menangkap hubungan nonlinier di data harga kendaraan bekas [16]. Selain itu, algoritma berbasis *boosting* seperti *XGBoost* juga terbukti mampu menghasilkan performa prediksi yang tinggi pada kasus serupa. Sistem prediksi berbasis kecerdasan buatan juga diketahui dapat meningkatkan akurasi dalam proses estimasi harga kendaraan [17].

Di sisi lain, model regresi linear masih dapat digunakan untuk prediksi harga kendaraan, namun memiliki keterbatasan dalam menangani hubungan antarvariabel yang

kompleks [13]. Oleh karena itu, metode berbasis *boosting* seperti *CatBoost* menjadi alternatif yang lebih efektif dibandingkan pendekatan konvensional.

Performa model yang baik dalam penelitian ini juga dipengaruhi oleh kemampuan *CatBoost* dalam menangani fitur kategorikal tanpa encoding yang rumit menjadikan *CatBoost* lebih sesuai untuk data kendaraan yang terdiri atas kombinasi atribut numerik dan kategorikal. Di samping itu, tahapan praproses seperti menangani data yang tidak lengkap, penghapusan data duplikat, serta transformasi atribut turut meningkatkan kualitas data sebelum proses pelatihan dilakukan. Penelitian sebelumnya juga menekankan bahwa kualitas data berpengaruh terhadap performa model prediksi harga kendaraan bekas [18].

Jika dibandingkan dengan pendekatan regresi linear berganda, model *CatBoost* menunjukkan keunggulan dalam menjelaskan variasi harga dan menjaga kestabilan performa antara data *training* dan *testing*. Regresi linear cenderung mengasumsikan hubungan linier antarvariabel sehingga kurang optimal apabila data pola tidak linier atau hubungan antar fitur yang bersifat kompleks. Temuan ini konsisten dengan studi sebelumnya yang mengungkapkan bahwa metode berbasis pohon keputusan dan *ensemble* memiliki akurasi yang lebih unggul dari pada regresi linear pada situasi tertentu [19],[20].

Dari sisi implementasi, penelitian ini menunjukkan bahwa model *CatBoost* tidak hanya memberikan performa prediksi yang tidak hanya baik, namun juga dapat diintegrasikan pada sistem berbasis web yang interaktif. Sistem memungkinkan pengguna memasukkan atribut kendaraan seperti merek, model, tahun produksi, kapasitas mesin, dan jarak tempuh untuk memperoleh estimasi harga secara cepat dan konsisten. Hasil penelitian ini mendukung temuan penelitian sebelumnya yang menyimpulkan bahwa sistem berbasis AI mampu membantu pengguna memperoleh estimasi harga kendaraan secara lebih objektif [17].

Selain menghasilkan prediksi harga, penelitian ini juga menunjukkan bahwa beberapa atribut memiliki pengaruh penting terhadap harga motor bekas, seperti model kendaraan, tahun produksi, umur kendaraan, kapasitas mesin, dan jarak tempuh. Temuan hal ini sejalan dengan studi terdahulu yang mengungkapkan bahwa atribut teknis kendaraan merupakan faktor utama dalam menentukan harga kendaraan bekas [12].

Meskipun dengan demikian, penelitian ini masih mempunyai sejumlah keterbatasan. Data yang digunakan hanya berasal dari satu platform, yaitu OLX Indonesia, sehingga belum sepenuhnya merepresentasikan kondisi pasar secara keseluruhan. Selain itu, variabel yang digunakan masih terbatas pada atribut yang tersedia pada dataset. Oleh karena itu, penelitian selanjutnya disarankan untuk menambahkan sumber data yang lebih beragam, memperluas variabel yang digunakan, serta membandingkan performa *CatBoost* dengan algoritma lainnya agar diperoleh model yang lebih optimal.

4. Simpulan

Penelitian ini berhasil mengimplementasikan algoritma Categorical Boosting (*CatBoost*) dalam sistem prediksi harga motor bekas berbasis web. Proses penelitian dilakukan melalui tahapan web scraping, prapemrosesan data, rekayasa fitur, pembagian dataset ke dalam data latih dan data uji untuk pelatihan model, hingga implementasi sistem dengan memanfaatkan framework Streamlit.

Berdasarkan hasil pengujian, algoritma *CatBoost* mampu menangani data kategorikal secara efektif dan menghasilkan performa prediksi yang baik pada data kendaraan bekas. Hasil evaluasi model menunjukkan bahwa nilai koefisien determinasi (R^2) yang diperoleh sebesar 0,9166 pada data latih dan 0,9149 pada data uji. Nilai tersebut mengindikasikan bahwa model memiliki kinerja yang baik dalam mengestimasi harga motor bekas sekaligus tidak menunjukkan indikasi overfitting yang signifikan.

Daftar Referensi

- [1] "Statistic Distribution – AISI," Asosiasi Industri Sepeda Motor Indonesia, 2026. [Online]. Available: <https://www.aisi.or.id/statistic/>. Accessed: Jan. 18, 2026.
- [2] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. 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.
- [3] 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.

- [4] M. R. Ardiansyah, Y. B. Utomo, and M. S. Anam, "Pengembangan Aplikasi Prediksi Harga Motor Bekas Berbasis Regresi Linear Berganda dengan Mempertimbangkan Faktor Penentu Harga," *JISCOM*, vol. 3, pp. 1–8, 2025.
- [5] M. A. Saputra, U. Hayati, T. Informatika, M. Informatika, R. Linier, and D. Mining, "Estimasi Harga Mobil Bekas Toyota Yaris Menggunakan Regresi Linear," *J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1696–1701, 2024.
- [6] 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.
- [7] L. P. Nasyuli, I. Lubis, and A. M. Elhanafi, "Penerapan Model Machine Learning Algoritma Gradient Boosting dan Linear Regression Melakukan Prediksi Harga Kendaraan Bekas," *J. Ilmu Komput. dan Sist. Inf.*, vol. 2, no. 2, pp. 299–310, 2023, doi: 10.70340/jirsi.v2i2.56.
- [8] A. S. B. A. F. Srg, "Sistem Informasi Peramalan Penjualan Sepeda Motor Bekas Menggunakan Metode Winter Exponential Smoothing pada CV Kenon Motor," *J. Rekayasa Sist.*, vol. 2, no. 3, pp. 1346–1359, 2024. [Online]. Available: <https://kti.potensi-utama.org/index.php/jureksi/article/view/1481>.
- [9] A. Masdiyanto, R. B. Kiyosaki, S. Hakiki, and others, "Application of the Naive Bayes Data Mining Algorithm to Predict Used Motorcycle Purchase Decisions," *J. Inform. Technol. Educ. Eng. Higher Appl. (JITEEHA)*, vol. 2, no. 1, 2025. [Online]. Available: <https://journal.iteeacademy.org/index.php/jiteeha/article/view/39>.
- [10] A. Professor, N. Niharika, S. Sravani, and D. Rakshini, "Enhanced Used Car Price Prediction Using Artificial Neural Networks and Ensemble Machine Learning Models," *J. Eng. Sci.*, vol. 16, no. 4, 2025.
- [11] U. Mishra, A. Rathi, A. Tomar, R. Saini, and S. Tomar, "Used Cars Price Prediction Using Machine Learning," *SSRN Electron. J.*, vol. 12, no. 4, pp. 970–976, 2024, doi: 10.2139/ssrn.4502021.
- [12] J. Huang, Z. Yu, Z. Ning, and D. Hu, *Used Car Price Prediction Analysis Based on Machine Learning*, vol. 1. Atlantis Press International BV, 2023, doi: 10.2991/978-94-6463-010-7_37.
- [13] M. D. H. Kusuma and S. Hidayat, "Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 2, pp. 1097–1110, 2024, doi: 10.35870/jmik.v5i2.629.
- [14] A. Faqih and T. Sugihartono, "Perbandingan Algoritma XGBoost dan LSTM dalam Prediksi Harga Saham Tesla Menggunakan Data Tahun 2025," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 6, pp. 1563–1573, 2025, doi: 10.52436/1.jpti.836.
- [15] P. A. Azhar, M. A. 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.
- [16] C. Longani, S. P. Potharaju, and S. Deore, "Price Prediction for Pre-owned Cars Using Ensemble Machine Learning Techniques," *Adv. Parallel Comput.*, vol. 39, pp. 178–187, 2021, doi: 10.3233/APC210194.
- [17] V. Waghmare, V. Hinchageri, P. Kadam, and B. B. Godbole, "AI-Based Vehicle Price Detection and Evaluation System," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 1557–1576, 2025, doi: 10.48175/ijarsct-29486.
- [18] J. Huang, "Price Prediction and Analysis of Price Influencing Factors for Second-hand Car Sales in AutoTrader Based on XGBoost Algorithm," *World Sci. Res. J.*, vol. 11, no. 9, 2025, doi: 10.6911/WSRJ.202509.
- [19] A. P. Kohan et al., "Komparasi Akurasi Regresi Linear Berganda dan Decision Tree Prediksi Harga Motor Bekas Berbasis Website," *INFOTECH J.*, vol. 11, no. 2, pp. 386–393, 2025, doi: 10.31949/infotech.v11i2.16430.
- [20] Y.R. Bere, & F.A.I Ahda, "Perbandingan Metode Decision Tree dan Logistic Regression dalam Klasifikasi Tingkat Obesitas Berdasarkan Gaya Hidup." *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 15, no. 2, pp. 746-756, 2026, doi:<http://dx.doi.org/10.35889/jutisi.v15i2.3591>.