JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 5, No. 1, Januari 2026, pp. 200 - 210
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/169

Implementation of the K-Nearest Neighbor Algorithm for Used Motorcycle Price Prediction Based on Vehicle Features

Muhamad Fahrul Rozi*[1], Lucky Heriyanto[2], Denny Rianditha Arief Permana[3], Ibnu Dzaky Abdad[4]
[1,2,3,4] Program Studi Sistem Informasi Industri Otomotif, Politeknik STMI Jakarta
Jl. Letjen Suprapto No.26, Cempaka Putih, Jakarta Pusat, DKI Jakarta, Indonesia
e-mail: fahrul@stmi.ac.id*[1], lucky@stmi.ac.id[2], dennyrian76@gmail.com[3], ibnuabdad14@gmail.com[4]

ARTICLE INFO

Article History:

Received: 01-11-205
Revised: 25-11-2025
Accepted: 01-12-2025
Published: 21-01-2026

ABSTRACT

The used motorcycle market in Indonesia shows an important trend as a viable mode of transportation. However, fair market price assessments are often subjective and pose a major challenge for sellers and buyers in obtaining accurate estimates. This study seeks to develop a predictive model for used motorcycle prices using the K-Nearest Neighbour (KNN) algorithm, based on historical data. The dataset used comprises 199 records, with attributes including model, year, transmission, odometer reading, vehicle type, taxation, fuel efficiency, and engine capacity. The research methodology includes data sanitation (addressing missing values through median imputation), pre-processing (combining One-Hot Encoding and Min-Max Scaling), and model construction with parameter setting k = 3. The findings reveal that the KNN model produced an R-squared (R^2) value of 0.9255, a Mean Absolute Error (MAE) of 816.55, and a Root Mean Squared Error (RMSE) of 1,027.20. These findings indicate that this model accounts for 92.55% of the variance in prices, making it a reliable instrument for decision support in the field of used motorcycle appraisal.

Keywords: *K-Nearest Neighbor, Price Prediction, Machine Learning, Used Motorcycle, Regression.*

Abstrak

Pasar sepeda motor bekas di Indonesia menunjukkan tren penting sebagai moda transportasi yang layak. Namun demikian, penilaian harga pasar yang adil sering ditandai oleh subjektivitas dan menimbulkan tantangan besar bagi penjual dan pembeli dalam mencapai estimasi yang akurat. Penelitian ini berusaha untuk mengembangkan model prediktif untuk harga sepeda motor bekas menggunakan algoritma *K-Nearest Neighbor* (KNN), berdasarkan data historis. *Dataset* yang digunakan terdiri dari 199 catatan dengan atribut yang mencakup model, tahun, transmisi, pembacaan odometer, jenis kendaraan, perpajakan, efisiensi bahan bakar, dan kapasitas mesin. Metodologi penelitian mencakup sanitasi data (mengatasi nilai yang hilang melalui imputasi median), pra-pemrosesan (menggabungkan *One-Hot Encoding* dan *Min-Max Scaling*), di samping konstruksi model dengan pengaturan parameter k = 3. Temuan mengungkapkan bahwa model KNN menghasilkan nilai R-kuadrat (R2) 0,9255, Kesalahan *Absolute* Rata-rata (MAE) 816,55, dan *Root Mean Squared Error* (RMSE) 1.027,20. Temuan ini menunjukkan bahwa model ini mahir dalam menghitung 92,55% dari *varians* dalam harga, menjadikannya instrumen yang dapat diandalkan untuk dukungan keputusan di bidang penilaian sepeda motor bekas.

Kata Kunci: *K-Nearest Neighbor, Prediksi Harga, Machine Learning, Motor Bekas, Regresi.*

DOI: https://doi.org/10.51903/c997er78

p-ISSN : 2803-1507 e-ISSN : 2803-1531

1. PENDAHULUAN

Pasar motor bekas di Indonesia telah menjadi pilihan utama bagi banyak konsumen, mengingat harganya yang lebih terjangkau dibandingkan motor baru. Namun, penentuan harga yang tepat untuk motor bekas tetap menjadi tantangan. Penjual sering kali kesulitan dalam menetapkan harga yang adil, yang dapat memengaruhi kelancaran transaksi. Faktor yang memengaruhi harga motor bekas mencakup merek, model, tahun produksi, kondisi fisik, odometer, jenis transmisi, dan kapasitas mesin [1]. Penggunaan metode prediktif berbasis data, seperti *Machine Learning* (ML), dapat membantu memitigasi subjektivitas dalam penetapan harga dan meningkatkan akurasi prediksi harga pasar. Salah satu algoritma ML yang dapat diterapkan adalah *K-Nearest Neighbor* (KNN), yang bekerja berdasarkan kedekatan data dalam ruang fitur, memungkinkan perhitungan harga motor bekas yang lebih objektif [2].

Saat ini, pasar motor bekas di Indonesia menunjukkan tren yang terus berkembang. Banyak konsumen lebih memilih membeli motor bekas untuk memenuhi kebutuhan transportasi harian mereka, terutama di kota-kota besar. Berdasarkan data yang diperoleh dari berbagai sumber [3], harga motor bekas dipengaruhi oleh faktor eksternal seperti kondisi pasar dan faktor internal seperti kondisi fisik kendaraan. Data yang ada menunjukkan bahwa motor bekas dengan jarak tempuh rendah dan tahun pembuatan yang lebih baru cenderung memiliki harga jual yang lebih tinggi. Fenomena ini mencerminkan kompleksitas dalam menentukan harga yang adil, yang mengarah pada potensi pemanfaatan *Business Intelligence* (BI) dan model prediksi berbasis ML untuk memberikan estimasi harga yang lebih tepat [4].

Penelitian terkait dengan prediksi harga motor bekas telah dilakukan sebelumnya dengan hasil yang cukup menjanjikan. *Metode Winter Exponential Smoothing* untuk memprediksi harga sepeda motor bekas dengan mempertimbangkan berbagai faktor seperti tahun, jarak tempuh, dan kapasitas mesin [5]. Penerapan regresi linear untuk prediksi harga mobil bekas dengan hasil yang juga menggambarkan relevansi algoritma untuk tugas serupa dalam industri otomotif. Penelitian ini berupaya untuk mengisi gap dengan fokus pada motor bekas di Indonesia, mengingat perbedaan yang signifikan dalam karakteristik pasar antara negara berkembang dan negara maju [6].

Berdasarkan studi literatur yang ada, terlihat adanya gap yang signifikan dalam hal penerapan algoritma KNN pada pasar motor bekas di Indonesia. Pada pasar mobil bekas atau menggunakan algoritma regresi dengan *Hyperparameter Tuning* untuk prediksi harga. Penelitian tentang motor bekas yang spesifik di Indonesia masih terbatas, dan belum ada yang secara khusus mengimplementasikan KNN untuk memprediksi harga motor bekas dengan mempertimbangkan variabel seperti kondisi fisik, pajak, dan konsumsi bahan bakar [7].

Penelitian ini bertujuan untuk menerapkan algoritma KNN dalam memprediksi harga motor bekas di Indonesia dengan menggunakan dataset yang mencakup berbagai atribut penting, seperti tahun pembuatan, odometer, tipe transmisi, dan kapasitas mesin. Selain itu, penelitian ini juga akan mengevaluasi akurasi model KNN dalam memprediksi harga motor bekas dengan menggunakan metrik *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE). Model ini diharapkan dapat memberikan estimasi harga yang lebih akurat, yang dapat digunakan oleh penjual dan pembeli untuk menetapkan harga yang lebih realistis dan berbasis data.

Sebagai Kesimpulan, penelitian ini diharapkan dapat mengembangkan model prediktif yang dapat diandalkan untuk pasar lokal. Selain itu, penelitian ini juga memberikan wawasan mengenai pengaruh berbagai faktor terhadap harga motor bekas yang mungkin belum banyak diteliti sebelumnya dalam konteks pasar Indonesia. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi baik dalam pengembangan ilmu pengetahuan di bidang *Business Intelligence* dan *Machine Learning*, maupun dalam praktik di industri otomotif untuk memfasilitasi transaksi jual beli motor bekas yang lebih transparan dan efisien.

2. TINJAUAN PUSTAKA

2.1. *Business Intelligence*

Dalam konteks penerapannya di perusahaan, *Business Intelligence* dimanfaatkan untuk menyediakan informasi historis yang mendukung analisis dan pelaporan guna menunjang manajemen serta pengambilan keputusan yang signifikan dalam meningkatkan efisiensi proses bisnis [8]. *Business Intelligence* (BI) memungkinkan organisasi tidak hanya untuk mengamankan daya saing, tetapi juga meningkatkan pendapatan dengan meningkatkan pengambilan keputusan [9]. Peneliti dan praktisi mendukung bahwa teknologi BI telah mencapai pertimbangan yang luar biasa dari keduanya organisasi dan industri [8].

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Pada lingkungan organisasi, *Business Intelligence* digunakan sebagai sistem atau aplikasi yang berperan dalam pengelolaan data guna menghasilkan informasi yang relevan [10].

2.2. *Data Mining*

Dalam berbagai kajian, penerapan data mining banyak digunakan untuk proses penggalian dan analisis data berukuran besar dengan memanfaatkan pendekatan statistika, kecerdasan buatan, dan *machine learning* guna mendukung perolehan informasi yang bernilai dalam berbagai bidang kajian [11], [12], [13]. Penelitian [14] menjelaskan penerapan data mining secara praktis melalui pemanfaatan perangkat lunak seperti *Orange* dan *RapidMiner* untuk mengolah ribuan opini acak di media sosial sehingga menghasilkan statistik sentimen yang terukur dan dapat digunakan dalam pengambilan keputusan. Dalam penelitian [15], penggunaan perangkat lunak seperti Orange dan RapidMiner dimanfaatkan untuk mengolah ribuan opini acak di media sosial menjadi statistik sentimen yang terukur dan berguna dalam pengambilan keputusan.

2.3. *Sistem Pendukung Keputusan*

Dalam penerapannya, Sistem Pendukung Keputusan (SPK) digunakan untuk menghasilkan alternatif berdasarkan kriteria tertentu guna membantu proses pemecahan masalah dan pengambilan keputusan [16]. Pada dasarnya SPK dirancang untuk mendukung seluruh tahap pengambilan keputusan mulai dari mengidentifikasi masalah, memilih data yang relevan, menentukan pendekatan yang digunakan dalam proses pengambilan keputusan, sampai mengevaluasi pemilihan alternatif [17]. SPK memanfaatkan sistem berbasis komputer dalam membantu pengambilan keputusan untuk memecahkan masalah tertentu melalui penggunaan data dan model [18]. Pada dasarnya SPK merupakan pengembangan lebih lanjut dari Sistem Informasi Manajemen terkomputerisasi yang dirancang sedemikian rupa sehingga bersifat interaktif dengan pemakainya [19].

2.4. *K-Nearest Neighbor (KNN)*

Dalam proses analisis data, metode *K-Nearest Neighbor* (KNN) digunakan untuk mengklasifikasikan objek dengan memanfaatkan data pelatihan yang memiliki jarak terdekat dengan objek yang dianalisis [20]. Metode K-Nearest Neighbor diterapkan untuk menemukan kasus yang paling relevan melalui pengukuran tingkat kesamaan antara kasus baru dan kasus yang telah ada dengan mempertimbangkan nilai serta bobot atribut [21]. Metode KNN digunakan dalam proses pengklasifikasian objek baru dengan memanfaatkan atribut serta sampel pelatihan [22]. Penelitian lain menjelaskan algoritma KNN adalah salah satu algoritma yang sudah popular. KNN ini termasuk ke dalam grup *instance based, learning*. Metode KNN merupakan merupakan teknik *lazy learning* [23].

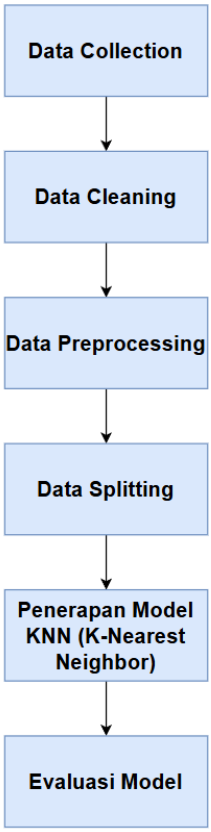
2.5. *Python*

Dalam penelitian dan pengembangan sistem, *Python* banyak dimanfaatkan untuk mendukung pengolahan data dan komputasi. *Python* juga didukung oleh pustaka (*library*) yang luas seperti *NumPy, Matplotlib,* dan *SymPy*, yang memungkinkan berbagai manipulasi matematika seperti kalkulus, aljabar simbolik, visualisasi grafik, hingga statistika komputasional [24]. Dalam praktiknya, eksekusi kode *Python* dilakukan melalui interpreter, sehingga mempermudah pengembangan dan pengujian program. Salah satu perbedaan utama *Python* dengan bahasa pemrograman lainnya adalah pengembangannya melibatkan jutaan *programmer*, peneliti, dan pengguna dari berbagai latar belakang, tidak hanya dari kalangan IT, karena *Python* bersifat *open source* [25].

3. METODOLOGI PENELITIAN

Metodologi penelitian ini dirancang untuk memberikan gambaran yang jelas mengenai tahapan yang dilakukan dalam penelitian prediksi harga motor bekas menggunakan algoritma *K-Nearest Neighbor* (KNN) [26]. Pendekatan yang digunakan adalah pendekatan kuantitatif dengan jenis penelitian eksperimen. Dalam penelitian ini, kami mengumpulkan dan menganalisis data historis yang terkait dengan motor bekas, serta menerapkan algoritma KNN untuk membangun model prediksi harga [27]. Berikut adalah tahapan dalam metodologi yang digunakan untuk penelitian ini mengadopsi pendekatan kuantitatif dengan metode eksperimen untuk membangun dan mengevaluasi model prediksi harga motor bekas menggunakan algoritma K-Nearest Neighbor (KNN) [28]. Langkah-langkah yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1.

p-ISSN : 2803-1507 e-ISSN : 2803-1531



Gambar 1. Tahapan Metodologi Penelitian

3.1. *Data Collection*

Pengumpulan data yang mencakup atribut-atribut penting motor bekas seperti model, tahun produksi, harga, jenis transmisi, *odometer*, jenis motor, pajak, konsumsi bahan bakar, dan kapasitas mesin. Data yang digunakan diambil dari *dataset* motor bekas yang berisi 199 entri dengan fitur-fitur yang relevan untuk prediksi harga motor bekas.

3.2. *Data Cleaning*

Setelah pengumpulan data, tahap berikutnya adalah pembersihan data. Pada tahap ini, dilakukan identifikasi terhadap missing values dan inkonsistensi dalam data. Data yang hilang pada kolom "Mesin" diatasi dengan menggunakan metode imputasi median untuk menggantikan nilai yang hilang dengan nilai tengah dari data yang ada. Proses konversi tipe data juga dilakukan untuk memastikan konsistensi dalam *format* data. Setelah itu, validasi data dilakukan untuk memastikan bahwa tidak ada lagi nilai yang hilang dan semua data telah dalam kondisi yang siap untuk diproses lebih lanjut.

3.3. *Data Preprocessing*

Selanjutnya, data yang telah dibersihkan menjalani tahap pra-pemrosesan, yang melibatkan pemilihan fitur dan target. Fitur numerik seperti tahun, *odometer*, pajak, konsumsi bahan bakar, dan kapasitas mesin akan dipilih untuk dimasukkan dalam model. Fitur kategorikal seperti transmisi dan jenis motor diubah menjadi format numerik menggunakan teknik *One-Hot Encoding* untuk memungkinkan model memprosesnya. Selain itu, normalisasi dilakukan dengan teknik *Min-Max Scaler* agar semua fitur berada dalam rentang yang sama, yang sangat penting karena KNN mengukur jarak antar data.

3.4. *Data Splitting*

Setelah tahap pra-pemrosesan, dataset dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Biasanya, sekitar 80% data digunakan untuk pelatihan model, sementara 20% sisanya digunakan untuk

p-ISSN : 2803-1507 e-ISSN : 2803-1531

menguji kemampuan model dalam memprediksi harga pada data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk memastikan bahwa model dapat diuji pada data yang representatif namun belum dikenali sebelumnya.

3.5. *Penerapan Model KNN*

Tahap berikutnya adalah penerapan model KNN. Model KNN dilatih menggunakan data pelatihan, di mana parameter k ditetapkan pada nilai 3, artinya model akan mempertimbangkan tiga tetangga terdekat dari data pengujian untuk memprediksi harga. KNN bekerja berdasarkan prinsip kesamaan atau kedekatan data, dan dalam hal ini, ia akan mencari motor bekas dengan karakteristik yang paling mirip untuk menghasilkan estimasi harga.

3.6. *Evaluasi Model*

Setelah model KNN dibangun, tahap terakhir adalah evaluasi model. Kinerja model diukur dengan menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-squared* (R^2). MAE mengukur rata-rata perbedaan absolut antara harga yang diprediksi dan harga aktual, sedangkan RMSE memberi bobot lebih pada kesalahan besar, dan R^2 mengukur sejauh mana variasi harga dapat dijelaskan oleh model. Metrik-metrik ini memberikan gambaran mengenai seberapa akurat model KNN dalam memprediksi harga motor bekas pada data pengujian.

4. HASIL DAN PEMBAHASAN

4.1. *Deskripsi Data Penelitian*

Dataset yang digunakan dalam penelitian ini berisi data terkait motor bekas yang dapat digunakan untuk memprediksi harga motor bekas berdasarkan beberapa variabel. Data ini terdiri dari 9 kolom, yang masing-masing menyimpan informasi yang relevan mengenai karakteristik motor bekas. Berikut adalah penjelasan mendetail mengenai kolom-kolom dalam *dataset* pada Tabel 1.

Tabel 1. Dataset Penelitian

Model	Tahun	Harga	Transmisi	Odometer	Jenis	Pajak	Konsumsi BBM	Mesin
Supra X 125	2017	7999	Manual	17307	Bebek	145	58.9	125
Vario 125	2016	14499	Automatic	25233	Skuter	235	43.5	125
Vario 125	2016	11399	Automatic	37877	Skuter	30	61.7	125
Revo	2016	6499	Manual	23789	Bebek	20	60.1	110
PCX	2015	10199	Automatic	33177	Skuter	160	51.4	150
Revo	2015	6149	Manual	41650	Bebek	20	60.1	110
Supra x 125	2015	6799	Manual	39547	Bebek	30	33.4	125
Vario 125	2016	14199	Automatic	30300	Skuter	125	57.6	125
Genio	2018	10399	Automatic	38082	Skuter	145	51.4	110
Vario 125	2016	12163	Automatic	21014	Skuter	30	61.7	125

1. Model: Kolom ini mencatat jenis atau model motor yang dimaksud. Setiap motor bekas dalam *dataset* ini memiliki model tertentu, yang dapat menunjukkan merk atau tipe motor tersebut. Model motor ini dapat memberikan gambaran mengenai kualitas dan keandalan motor berdasarkan tipe yang ada di pasaran.
2. Tahun: Kolom ini menunjukkan tahun pembuatan motor. Tahun produksi ini sangat penting karena motor dengan tahun pembuatan yang lebih baru biasanya memiliki harga yang lebih tinggi, mengingat umur motor yang lebih muda dan kemungkinan perbaikan yang lebih sedikit. Di sisi lain, motor dengan tahun produksi lebih lama biasanya memiliki harga lebih rendah karena faktor depresiasi.
3. Harga: Kolom ini berisi harga jual motor bekas yang menjadi target prediksi dalam penelitian ini. Harga ini dipengaruhi oleh berbagai faktor seperti tahun pembuatan, model motor, jumlah kilometer yang sudah ditempuh, dan beberapa variabel lainnya. Ini adalah variabel dependen yang akan diprediksi menggunakan model yang dibangun dalam penelitian ini.
4. Transmisi: Kolom ini menunjukkan jenis transmisi motor, apakah motor tersebut menggunakan transmisi manual atau otomatis. Jenis transmisi ini mempengaruhi kenyamanan dan kemudahan penggunaan motor, yang pada gilirannya juga memengaruhi harga jual motor bekas. Motor dengan transmisi otomatis biasanya memiliki harga yang lebih tinggi dibandingkan dengan motor transmisi manual.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

5. *Odometer*: Kolom ini menunjukkan jumlah kilometer yang telah ditempuh oleh motor. Angka odometer ini menjadi indikator seberapa sering motor digunakan, yang secara langsung mempengaruhi kondisi dan harga motor bekas. Motor dengan jarak tempuh yang rendah biasanya dihargai lebih tinggi karena dianggap masih memiliki umur pakai yang lebih lama.
6. Jenis: Kolom ini mencatat jenis motor, apakah motor tersebut termasuk dalam kategori bebek, skuter, sport, atau jenis lainnya. Jenis motor ini berperan penting dalam menentukan harga pasar motor bekas. Motor sport, misalnya, cenderung lebih mahal daripada motor bebek atau skuter, karena motor sport memiliki performa yang lebih tinggi.
7. Pajak: Kolom ini mencatat jumlah pajak kendaraan yang telah dibayarkan. Pajak kendaraan bermotor merupakan salah satu faktor yang dapat memengaruhi harga jual motor bekas, karena motor yang memiliki pajak yang terbayar dengan lengkap dan tepat waktu lebih dihargai oleh pembeli potensial.
8. Konsumsi BBM: Kolom ini menunjukkan konsumsi bahan bakar per kilometer yang diperlukan oleh motor. Faktor efisiensi bahan bakar ini dapat mempengaruhi daya tarik motor bekas di pasar, karena motor yang lebih irit bahan bakar cenderung lebih diminati oleh pembeli yang peduli dengan biaya operasional motor dalam jangka panjang.
9. Mesin: Kolom ini menunjukkan ukuran mesin motor dalam satuan cc (*centimeter* kubik). Ukuran mesin ini sangat berpengaruh terhadap performa motor dan harga jual motor bekas. Motor dengan mesin yang lebih besar cenderung memiliki harga yang lebih tinggi, karena menawarkan performa yang lebih kuat.

4.2. *Data Cleaning*

Proses pembersihan data merupakan langkah awal yang sangat penting dalam analisis data karena data yang tidak bersih dapat mengarah pada hasil analisis yang tidak akurat. Dalam penelitian ini, data yang digunakan berasal dari dataset motor bekas yang terdiri dari 199 entri dan 9 atribut. Setelah dilakukan pengecekan terhadap data, ditemukan bahwa terdapat nilai yang hilang (*missing values*) pada kolom mesin. Untuk menangani masalah ini, dilakukan pengisian nilai yang hilang dengan menggunakan nilai median dari kolom mesin tersebut. Penggunaan median dipilih karena nilai median tidak dipengaruhi oleh data yang ekstrem, sehingga dapat memberikan nilai yang lebih representatif bagi data yang hilang.

```
# 2. Atasi Missing Value pada kolom 'mesin'
# Hitung median dari kolom 'mesin'
median_mesin = df['mesin'].median()

# Isi missing value dengan median
df['mesin'] = df['mesin'].fillna(median_mesin)
```

Setelah melakukan pengisian, kolom mesin kemudian diubah tipe datanya dari *float*64 menjadi integer agar konsisten dengan tipe data lainnya yang bersifat numerik.

```
# Konversi tipe data kolom 'mesin' ke integer
df['mesin'] = df['mesin'].astype(int)
```

Setelah proses pembersihan selesai, dilakukan pengecekan kembali untuk memastikan bahwa seluruh data sudah dalam keadaan *non-null*. Hasilnya, dataset kini memiliki 199 entri yang lengkap tanpa adanya nilai yang hilang. Ini menunjukkan bahwa data sudah siap untuk digunakan dalam tahap analisis berikutnya. Hasil dari proses pembersihan data ini dapat dilihat pada Gambar 2.

```
# Cek kembali untuk memastikan semua non-null
print("\n--- Hasil Pembersihan Data ---")
df.info()
```

p-ISSN : 2803-1507 e-ISSN : 2803-1531

```
--- Hasil Pembersihan Data ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 199 entries, 0 to 198
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   model        199 non-null    object
 1   tahun        199 non-null    int64
 2   harga        199 non-null    int64
 3   transmisi    199 non-null    object
 4   odometer     199 non-null    int64
 5   jenis        199 non-null    object
 6   pajak        199 non-null    float64
 7   konsumsiBBM  199 non-null    float64
 8   mesin        199 non-null    int64
dtypes: float64(2), int64(4), object(3)
memory usage: 14.1+ KB
```

Gambar 2. Hasil Data Cleaning

4.3. *Data Preprocessing*

Setelah dilakukan pembersihan data, langkah selanjutnya adalah *preprocessing* untuk mempersiapkan data agar dapat digunakan dalam model prediksi. Beberapa tahapan *preprocessing* yang dilakukan adalah:

1. *Encoding*: Kolom transmisi dan jenis yang merupakan data kategorikal diubah menjadi format numerik menggunakan metode *One-Hot Encoding*. Teknik ini digunakan untuk menghindari pengaruh urutan pada data kategorikal yang tidak bersifat numerik. Sebagai contoh, kolom transmisi yang memiliki dua kategori yaitu "Manual" dan "Automatic" diubah menjadi dua kolom baru dengan nilai 1 atau 0, yang menunjukkan apakah motor tersebut menggunakan transmisi manual atau otomatis.

```
df_encoded = pd.get_dummies(df, columns=['transmisi', 'jenis'],
drop_first=True)
```

2. Pemilihan Fitur dan Target

```
# 3. Pemilihan Fitur dan Target
# Kita hanya menggunakan fitur numerik untuk penyederhanaan awal model KNN
features_numerik = ['tahun', 'odometer', 'pajak', 'konsumsiBBM', 'mesin']
features_kategorikal_encoded = [col for col in df_encoded.columns if
col.startswith(('transmisi_', 'jenis_'))]
# Gabungkan semua fitur
all_features = features_numerik + features_kategorikal_encoded

X = df_encoded[all_features] # Fitur
Y = df_encoded['harga']      # Target
```

3. Normalisasi: Semua fitur numerik dalam dataset dinormalisasi menggunakan *Min-Max Scaler*. Normalisasi dilakukan untuk memastikan bahwa semua fitur berada dalam rentang yang sama (0 hingga 1). Hal ini sangat penting karena algoritma KNN menghitung jarak antara data, dan jika fitur memiliki rentang yang berbeda jauh, fitur dengan nilai besar dapat mendominasi hasil perhitungan jarak.

```
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
X_scaled_df = pd.DataFrame(X_scaled, columns=all_features)
```

4.4. *Data Split*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Setelah proses encoding dan normalisasi, data dibagi menjadi dua set, yaitu data pelatihan dan data pengujian. Pembagian dilakukan dengan menggunakan fungsi `train_test_split` dari pustaka scikit-learn, yang membagi data secara acak dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Hal ini bertujuan agar model dapat dilatih dengan sebagian besar data dan diuji pada data yang belum pernah dilihat sebelumnya, sehingga dapat mengukur kemampuannya dalam memprediksi harga motor bekas pada data yang baru.

```
# 5. Membagi Data Training dan Testing
# Menggunakan 80% data untuk training dan 20% untuk testing
X_train, X_test, Y_train, Y_test = train_test_split(
    X_scaled_df, Y, test_size=0.2, random_state=42
)

print(f"\nJumlah data Training: {len(X_train)} baris")
print(f"Jumlah data Testing: {len(X_test)} baris")
```

4.5. *Implementasi Model KNN*

Dalam implementasi model KNN ini, parameter yang digunakan adalah nilai k=3, yang berarti bahwa untuk memprediksi harga motor bekas, model akan mempertimbangkan tiga tetangga terdekat dari data pengujian. di mana prediksi berdasarkan rata-rata hasil dari k data paling mirip dengan data yang ingin diestimasi. Selanjutnya, objek `model_knn` dibuat dari kelas *KNeighborsRegressor* dan kemudian dilatih dengan data training menggunakan fungsi fit, sehingga model dapat mengetahui pola hubungan antar fitur dan target, serta siap digunakan untuk melakukan prediksi nilai target (`Y_pred`) pada data testing yang belum pernah dilihat sebelumnya.

```
# 6. Implementasi Model KNN
# Kita coba dengan k=3 (3 tetangga terdekat)
k = 3
model_knn = KNeighborsRegressor(n_neighbors=k)

# Melatih model dengan data training
model_knn.fit(X_train, Y_train)

# 7. Prediksi
Y_pred = model_knn.predict(X_test)
```

4.6. *Evaluasi Model*

Untuk menilai kinerja model, dilakukan evaluasi dengan menggunakan tiga metrik utama, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-squared* (R^2).

1. *Mean Absolute Error* (MAE): Metrik ini mengukur rata-rata perbedaan absolut antara harga yang diprediksi oleh model dan harga aktual motor bekas. Nilai MAE yang lebih kecil menunjukkan prediksi yang lebih akurat. Berdasarkan hasil evaluasi, nilai MAE yang diperoleh adalah 816.55, yang berarti prediksi harga motor bekas memiliki kesalahan rata-rata sekitar 816 ribu rupiah.
2. *Root Mean Squared Error* (RMSE): RMSE mengukur akar kuadrat dari rata-rata kuadrat kesalahan antara harga aktual dan harga prediksi. Nilai RMSE yang diperoleh adalah 1,027.20, yang menunjukkan bahwa model ini masih memiliki kesalahan dalam memprediksi harga motor bekas, terutama pada motor dengan harga yang lebih tinggi.
3. *R-squared* (R^2): R^2 mengukur sejauh mana variasi harga motor bekas dapat dijelaskan oleh model. Nilai R^2 yang diperoleh adalah 0.9255, yang berarti model ini mampu menjelaskan sekitar 92.55% variasi harga motor bekas dalam dataset. Semakin mendekati nilai 1, semakin baik model dalam menjelaskan variasi data.

```
# 8. Evaluasi Model
mae = mean_absolute_error(Y_test, Y_pred)
mse = mean_squared_error(Y_test, Y_pred)
rmse = np.sqrt(mse)
```

p-ISSN : 2803-1507 e-ISSN : 2803-1531

```
r2 = r2_score(Y_test, Y_pred)

print(f"\n--- Hasil Evaluasi Model KNN (k={k}) ---")
print(f"Mean Absolute Error (MAE): {mae:,.2f}")
print(f"Root Mean Squared Error (RMSE): {rmse:,.2f}")
print(f"R-squared (R2 Score): {r2:.4f}")

# Tampilkan perbandingan Harga Aktual vs Prediksi
hasil_prediksi = pd.DataFrame({
    'Harga Aktual': Y_test,
    'Harga Prediksi': Y_pred.round(2)
})

print("\n--- Perbandingan Harga Aktual vs. Prediksi (Top 5) ---")
print(hasil_prediksi.head().to_markdown(numalign="left", stralign="left"))
```

4.7. *Perbandingan Harga Aktual dan Prediksi*

Sebagai bagian dari evaluasi, perbandingan antara harga aktual dan harga yang diprediksi oleh model dilakukan pada data pengujian. Analisis ini dilakukan untuk menilai seberapa akurat model KNN dalam memprediksi harga motor bekas. Dari lima data pertama, terlihat bahwa prediksi model cukup mendekati harga aktual, meskipun terdapat beberapa perbedaan kecil. Hasil perbandingan antara harga aktual dan harga prediksi dapat dilihat pada Gambar 3.

```
--- Hasil Evaluasi Model KNN (k=3) ---
Mean Absolute Error (MAE): 816.55
Root Mean Squared Error (RMSE): 1,027.20
R-squared (R2 Score): 0.9255

--- Perbandingan Harga Aktual vs. Prediksi (Top 5) ---
|     | Harga Aktual   | Harga Prediksi   |
|:----|:---------------|:-----------------|
| 82  | 8000           | 7632.33          |
| 15  | 8700           | 10465.3          |
| 111 | 15400          | 16516            |
| 177 | 15298          | 15065.3          |
| 76  | 15700          | 15532            |
```

Gambar 3. Hasil Evaluasi Model KNN

Hal ini menunjukkan bahwa model KNN dapat memprediksi harga motor bekas dengan cukup akurat, meskipun terdapat beberapa perbedaan antara harga yang diprediksi dan harga aktual. Meskipun tidak selalu tepat, perbedaan harga prediksi dan aktual tidak terlalu besar, sehingga model ini dapat digunakan untuk memprediksi harga motor bekas dengan tingkat akurasi yang cukup baik.

5. DISKUSI

Hasil pengujian menunjukkan bahwa model *K-Nearest Neighbor* dengan parameter k = 3 mampu menjelaskan 92,55% variasi harga motor bekas, sebagaimana tercermin dari nilai R^2 sebesar 0,9255. Nilai ini menunjukkan bahwa kombinasi variabel tahun produksi, odometer, pajak, konsumsi bahan bakar, kapasitas mesin, serta fitur kategorikal yang telah melalui *one-hot encoding* memiliki daya representasi yang kuat terhadap pola harga pasar. Tingkat kesalahan prediksi yang ditunjukkan oleh MAE sebesar 816,55 dan RMSE sebesar 1.027,20 mengindikasikan adanya deviasi prediksi yang relatif kecil terhadap harga aktual. Pola ini memperlihatkan bahwa pendekatan berbasis jarak pada *feature space* mampu menangkap kemiripan karakteristik kendaraan secara efektif.

Jika dibandingkan dengan penelitian [2] yang menerapkan *KNN regressor* pada prediksi harga mobil bekas, nilai R^2 yang diperoleh pada penelitian ini berada pada tingkat yang sebanding meskipun objek penelitian berbeda. Penelitian [3] yang menggunakan algoritma C4.5 menunjukkan bahwa variabel jarak tempuh dan tahun produksi juga menjadi penentu utama harga, sejalan dengan temuan dalam penelitian ini. Namun

p-ISSN : 2803-1507 e-ISSN : 2803-1531

demikian, penelitian [6] yang menerapkan regresi linear pada mobil bekas melaporkan nilai akurasi yang lebih rendah, yang mengindikasikan keterbatasan model linier dalam menangkap hubungan nonlinier antarvariabel. Perbandingan ini menunjukkan bahwa *KNN* memiliki keunggulan dalam konteks pasar kendaraan bekas yang bersifat heterogen.

Analisis kesalahan prediksi memperlihatkan bahwa deviasi terbesar terjadi pada motor dengan harga tinggi, terutama pada segmen skuter otomatis dengan kapasitas mesin besar. Fenomena ini berkaitan dengan sensitivitas *KNN* terhadap distribusi data lokal, di mana jumlah tetangga dengan karakteristik serupa pada rentang harga tinggi relatif lebih sedikit. Kondisi ini juga dilaporkan oleh [7] dalam studi prediksi harga kendaraan yang menekankan pentingnya kepadatan data pada setiap segmen harga. Dengan demikian, variasi distribusi data berperan langsung terhadap stabilitas hasil prediksi pada algoritma berbasis jarak.

Dari sisi implikasi, hasil penelitian ini menunjukkan bahwa model *KNN* dapat dimanfaatkan sebagai komponen *decision support system* dalam penentuan harga motor bekas berbasis data. Secara teoretis, penelitian ini memperluas penerapan *machine learning regression* pada konteks pasar motor bekas di Indonesia yang sebelumnya masih terbatas pada mobil bekas atau metode peramalan konvensional [5]. Secara praktis, integrasi variabel teknis kendaraan dan atribut administratif seperti pajak memberikan gambaran harga yang lebih mendekati kondisi pasar aktual. Namun, interpretasi hasil tetap perlu mempertimbangkan keterbatasan jumlah data dan ketergantungan *KNN* terhadap pemilihan parameter k serta distribusi data latih.

6. KESIMPULAN DAN SARAN

6.1. *Kesimpulan*

Penelitian ini berhasil mengimplementasikan algoritma *K-Nearest Neighbor* (KNN) untuk memprediksi harga motor bekas dengan hasil yang memuaskan. Model KNN dengan parameter k=3 mampu memberikan prediksi harga yang akurat, dengan skor R^2 mencapai 0,9255, yang menunjukkan bahwa model ini dapat menjelaskan sebagian besar variasi harga motor bekas. Tingkat kesalahan prediksi, yang diukur dengan *Mean Absolute Error* (MAE) sebesar 816,55, menunjukkan bahwa model ini cukup presisi untuk dijadikan acuan dalam estimasi harga pasar. Selain itu, tahapan preprocessing data, seperti penanganan missing value dan normalisasi data, memiliki peran yang sangat penting dalam meningkatkan akurasi model KNN yang berbasis jarak, membuktikan bahwa pemrosesan data yang tepat dapat memberikan dampak besar terhadap kinerja algoritma.

6.2. *Saran*

1. Penambahan Dataset: Jumlah data yang digunakan dalam penelitian ini terbatas pada 199 baris. Oleh karena itu, disarankan untuk menambah jumlah data agar model lebih robust dan mampu mengakomodasi variasi motor bekas yang lebih luas.
2. Hyperparameter Tuning: Diperlukan pencarian nilai k yang lebih optimal dengan menggunakan teknik seperti *GridSearch* CV Ini bertujuan untuk memastikan bahwa nilai k=3 adalah yang paling sesuai atau menemukan nilai lain yang lebih optimal dalam memprediksi harga motor bekas.
3. Perbandingan dengan Algoritma Lain: Disarankan untuk membandingkan hasil KNN dengan algoritma lain, seperti *Random Forest Regressor* atau Linear Regression, guna mendapatkan gambaran yang lebih komprehensif terkait efisiensi komputasi dan akurasi model dalam memprediksi harga motor bekas.

DAFTAR PUSTAKA

[1] S. Gunawan, "Analisis Sadd Adz-Dzari'ah terhadap Praktik Jual Beli Sepeda Motor Bekas Tanpa Surat: Studi Kasus Kabupaten Mandailing Natal Sumatera Utara," *Indones. J. Humanit. Soc. Sci.*, vol. 5, no. 4, pp. 2029–2050, 2024, doi: 10.33367/ijhass.v5i4.6427.

[2] M. S. Ghaisani and A. Baita, "Implementation of the K-Nearest Neighbors (KNN) Regressor Method to Predict Toyota Used Car Prices," *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 194–201, 2025, doi: 10.30871/jaic.v9i1.8860.

[3] S. Mulyanda, S. Defit, and Sumijan, "Analisis Data Mining Menggunakan Algoritma C4.5 untuk Prediksi Harga Pasar Mobil Bekas," *J. KomtekInfo*, vol. 10, no. 3, pp. 116–121, 2023, doi: 10.35134/komtekinfo.v10i3.427.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[4] I. D. Komang, D. Pramartha, G. Made, and A. Sasmita, "Penerapan Business Intelligence untuk Prediksi Penjualan Produk (Studi Kasus PT. XYZ)," *JITTER J. Ilm. Teknol. dan Komput.*, vol. 4, no. 2, pp. 1868–1879, 2023, doi: 10.24843/jtrti.2023.v04.i02.p19.

[5] A. S. b Ade Fitriani 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

[6] E. Hasibuan, "Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web," *J. Ilm. Komputasi*, vol. 21, no. 4, pp. 595–602, 2022, doi: 10.32409/jikstik.21.4.3327.

[7] A. Amalia, M. Radhi, S. H. Sinurat, D. R. H. Sitompul, and E. Indra, "Prediksi Harga Mobil Menggunakan Algoritma Regressi Dengan Hyper-Parameter Tuning," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 4, no. 2, pp. 28–32, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2479.

[8] S. Maesaroh, R. R. Lubis, L. N. Husna, R. Widyaningsih, and R. Susilawati, "Efektivitas Implementasi Manajemen Business Intelligence pada Industri 4.0," *ADI Bisnis Digit. Interdisiplin J.*, vol. 3, no. 2, pp. 1–8, 2022, doi: 10.34306/abdi.v3i2.764.

[9] A. W. Nugroho and A. A. G. S. Utama, "Business Intelligence Systems and Their Impact on Organizational Decision-Making and Performance Outcomes: Literature Review," *Owner*, vol. 9, no. 2, pp. 1269–1284, 2025, doi: 10.33395/owner.v9i2.2646.

[10] A. R. Nurridwan Firdaus and D. Firmansyah, "Implementasi Business Intelligence pada Data Pendapatan studi kasus (PT. Pos Indonesia)," *J. Esensi Infokom J. Esensi Sist. Inf. dan Sist. Komput.*, vol. 7, no. 2, pp. 33–39, 2023, doi: 10.55886/infokom.v7i2.686.

[11] C. Supriadi, C. Sitinjak, D. Fujiama, N. D. Silalahi, S. Pertiwi, and A. Umar, "Enhancing Performance Using New Hybrid Intrusion Detection System," *J. Technol. Informatics Eng.*, vol. 1, no. 2, pp. 1–4, Jul. 2022, doi: 10.51903/jtie.v1i2.134.

[12] M. H. Haiedar and S. Kholifah, "The Impact of Occupational Safety, Workload, and Compensation on Employee Job Satisfaction in the Mining Industry: A Case Study of PT BCKA," *J. Manag. Informatics*, vol. 4, no. 1, pp. 587–598, Jan. 2025, doi: 10.51903/jmi.v4i1.53.

[13] F. G. Falentina, G. Y. Y. Wabdaron, D. Andiyani, M. S. Wondiwoi, and H. Sutejo, "Analisis Penerapan Data Mining Untuk Klasifikasi Penjualan Makanan Terlaris Menggunakan Algoritma Decision Tree (C4.5)," *J. Ilm. Sist. Inf.*, vol. 4, no. 2, pp. 382–395, Jan. 2025, doi: 10.51903/f2jegk76.

[14] T. Y. Pahtoni and H. Jati, "Analisis Sentimen Data Twitter Terkait Chatgpt Menggunakan Orange Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 2, pp. 329–336, 2024, doi: 10.25126/jtiik.20241127276.

[15] E. Kurnia, *Penerapan Data Mining Menggunakan Algoritma Apriori dan Metode K-Nearest Neighbor dalam Menentukan Persediaan Barang Sembako UD. Amora Jaya*, 1st ed. Malang: PT. Literasi Nusantara Abadi Grup, 2025. [Online]. Available: https://repository-penerbitlitnus.co.id/id/eprint/495/1/PENERAPAN DATA MINING.pdf

[16] M. Yanto, "Sistem Penunjang Keputusan Dengan Menggunakan Metode AHP Dalam Seleksi Produk," *J. Teknol. dan Sist. Inf. Bisnis*, vol. 3, no. 1, pp. 167–174, 2021, doi: 10.47233/jteksis.v3i1.161.

[17] H. Hertyana, E. Mufida, and A. Al Kaafi, "Sistem Pendukung Keputusan Pemilihan Laptop dengan Menggunakan Metode Topsis," *J. Tek. Inform. UNIKA St. Thomas*, vol. 5, no. 1, pp. 36–44, 2021, doi: 10.54367/jtiust.v6i1.1216.

[18] B. Aji, D. Sasmoko, and C. Supriadi, "Sistem Pendukung Keputusan Penentuan Pembelian Laptop Kualitas Terbaik di Guava Komputer dengan Metode AHP Berbasis Web," *J. Ilm. Sist. Inf.*, vol. 3, no. 3, pp. 20–29, 2024, doi: 10.51903/avceqq43.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[19] Aldo, S. Muhammad Arifin, and W. Sri, "Sistem Pendukung Keputusan Penilaian Kinerja Dosen Dengan Menggunakan Metode Multi Attribute Utility Theory (MAUT)," *J. Sist. Inf. dan Manaj.*, vol. 7, no. 2, pp. 15–23, 2019, [Online]. Available: https://id.scribd.com/document/948993757/194-article-text-716-1-10-20200625

[20] W. R. Hasibuan, I. P. Sari, and M. Basri, "Klasifikasi Kerusakan (Cacat) pada Biji Kopi Arabika Menggunakan Algoritma KNN (K-Nearest Neighbor)," *Blend Sains J. Tek.*, vol. 3, no. 4, pp. 452–459, 2025, doi: 10.56211/blendsains.v3i4.781.

[21] T. Prasetyo, A. A. Waskita, and T. Taryo, "Analisis Sentimen Pengguna Seputar Kendaraan Listrik di Twitter dengan Penerapan Algoritma Naïve Bayes, KNN, dan Decision Tree untuk Klasifikasi," *J. SISKOM-KB*, vol. 8, no. 2, pp. 108–117, 2025, doi: 10.47970/siskom-kb.v8i1.783.

[22] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, no. 1, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.

[23] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes terhadap Analisis Sentimen Pengguna Layanan BPJS," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021, doi: 10.32493/informatika.v5i4.7622.

[24] S. Suwanto, N. Zega, A. Fitri, and M. Siddik, "Implementasi Softwere Python terhadap Pembelajaran Matematika Tingkat SMA: Studi Literatur," *Invent. J. Res. Educ. Stud.*, vol. 6, no. 3, pp. 306–314, 2025, doi: 10.51178/invention.v6i1.2537.

[25] S. Juni Anggita, Y. Yahfizham, and H. Habib Syarkowi, "Implementasi HTML dan JavaScript dalam Media Pembelajaran Matematika Berbasis Web: A Systematic Literature Review," *Math. Data Anal.*, vol. 2, no. 1, pp. 152–163, 2025, doi: 10.47709/mda.v2i01.6227.

[26] R. Hidayatullah, D. A. Fatah, and A. Yasid, "Penerapan Algoritma K-Nearest Neighbor pada Minat Beli Mobil Bekas Menggunakan Pendekatan CRISP-DM," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 2210–2217, 2025, doi: 10.36040/jati.v9i2.13024.

[27] M. Sulaeman, Y. A. Wijaya, and S. Sapriadi, "Klasifikasi Kondisi Fisik Mobil Bekas Menggunakan Algoritma Neural Network (NN)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1555–1559, 2024, doi: 10.36040/jati.v8i2.9020.

[28] G. Kelvin and A. Rudy, "Penerapan Data Mining untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor pada CV. Nuansa Karuna Lestari," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 6, pp. 11184–11190, 2024, doi: 10.36040/jati.v8i6.11189.