

Implementation of the K-Nearest Neighbors (KNN) Regressor Method to Predict Toyota Used Car Prices

Mauhiba Salmaa Ghaisani ^{1*}, Anna Baita ^{2*}

* Informatika, Universitas AMIKOM Yogyakarta
salmaa03@students.amikom.ac.id ¹, anna@amikom.ac.id ²

Article Info

Article history:

Received 2024-11-10

Revised 2024-12-03

Accepted 2025-01-18

Keyword:

*KNN Regressor,
Price Prediction,
Toyota Used Cars.*

ABSTRACT

The development of the automotive industry in Indonesia has experienced significant growth in recent decades, especially in the used car market segment. One of the used car brands that has high demand is Toyota, because it has a reliable reputation and quality. However, there are challenges that are often faced by sellers and buyers of used cars, namely in determining prices correctly and accurately. Incorrect pricing can be detrimental to one party, either the price is too high or too low. Prices that are too high can slow down the turnover of goods in the market. While low prices can cause sellers to experience losses. The purpose of this study is to help find good performance in determining the price of used Toyota cars. This study will use one of the Machine Learning methods, namely K-Nearest Neighbors Regressor. The KNN method is one method that can be used for classification and regression. In addition, this algorithm is a simple algorithm and can provide accurate prediction results based on its proximity to existing data. This study uses selected relevant features, namely model, year, kilometer, tax, mpg, and cc. The results of this study obtained MAE = 3.31686, MSE = 26.43640, RMSE = 5.14163, and R2-Score = 0.99501 using 90:10 data division and k = 1. This proves that KNN Regressor is an effective method in predicting the price of used Toyota cars. Therefore, the K-Nearest Neighbors (KNN) Regressor method is able to provide a fairly accurate price estimate with a minimal error rate.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Perkembangan bisnis di Indonesia telah mengalami kemajuan yang pesat, terutama dalam sektor industri otomotif. Hal ini dapat dilihat dari banyaknya jenis dan merek mobil yang dikeluarkan di Indonesia [1]. Kendaraan mobil adalah salah satu jenis transportasi yang banyak digunakan oleh masyarakat. Keberadaan mobil sangat mendukung aktivitas manusia dalam menjalankan berbagai kegiatan sehari-hari [2]. Salah satu contohnya yaitu mobil merek Toyota. Toyota merupakan salah satu merek mobil yang masih sering dicari dan merek terlaris dipasaran mencapai 304.736 unit [3]. Akan tetapi harga kendaraan mobil yang masih terbilang mahal menjadikan alasan masyarakat memilih untuk membeli mobil bekas. Peningkatan berbagai fitur pada mobil baru menjadi penyebab kenaikan harga mobil

secara signifikan [4]. Semakin banyak inovasi baru yang diciptakan, membuat harga mobil bekas semakin murah. Oleh karena itu, masyarakat lebih memilih untuk membeli mobil bekas agar mendapat harga yang lebih terjangkau. Banyaknya minat masyarakat terhadap mobil bekas menjadikan peluang untuk para pebisnis [5].

Dalam proses jual beli mobil bekas, terdapat tantangan yang sering dihadapi oleh penjual maupun pembeli yaitu dalam menentukan harga mobil dengan tepat dan akurat. Harga yang cukup tinggi dapat mempersulit proses penjualan, sehingga dapat memperlambat perputaran barang yang ada di pasar. Sedangkan harga yang rendah dapat membuat penjual mengalami kerugian, dikarenakan harga kendaraan tidak sesuai dengan kondisi dan spesifikasi dari mobilnya [6]. Dalam kasus ini, prediksi harga menjadi solusi dalam menetapkan harga [7]. Prediksi harga yang tepat dan akurat

sangat penting agar proses jual beli dapat berjalan secara adil dan transparan. Selain itu terdapat beberapa faktor-faktor yang kompleks yang mempengaruhi harga mobil bekas, seperti tahun pembuatan, jarak tempuh, model, dan kondisi fisik dari kendaraan. Informasi tentang harga mobil bekas sangat penting dan dibutuhkan untuk membantu masyarakat mengetahui estimasi harga mobil. Oleh karena itu, dalam menentukan harga yang realistis diperlukan metode yang dapat menganalisis faktor-faktor yang mempengaruhi harga mobil bekas dengan baik [8].

Machine learning merupakan cabang dari *Artificial Intelligence* (AI) yang dapat digunakan untuk menghasilkan model yang nantinya dapat menghasilkan prediksi [9]. Salah satunya memprediksi harga. Dalam melakukan prediksi harga, model *machine learning* yang sering digunakan yaitu Linear Regression [7], [10], [11], *Random Forest* [11], [12], *Long Short-Term Memory* [13], *Decision Tree* [12], dan *K-Nearest Neighbor* [14], [15], [16]. Algoritma *K-Nearest Neighbor* (KNN) dinilai dapat melakukan prediksi harga dengan baik [16]. KNN memerlukan nilai k dan matrik k jarak dengan jumlah yang relatif kecil jika dibandingkan dengan algoritma lainnya [17]. Selain itu KNN juga merupakan algoritma yang cukup sederhana. Ketika sampel training baru ditambahkan, KNN secara otomatis akan memperhitungkan data baru tersebut karena seluruh data pelatihan disimpan dalam memori [16].

Penelitian dengan menerapkan algoritma *K-Nearest Neighbors* (KNN) telah banyak dilakukan sebelumnya, di antaranya penelitian [16], [15], [18]. Penelitian Mukhlisin, dkk melakukan penelitian prediksi harga beras premium dengan membandingkan nilai RMSE menggunakan beberapa model *machine learning* seperti *K-Nearest Neighbor* (KNN) *Regressor*, *Support Vector Regression* (SVR), *Multilayer Perceptron* (MLP), dan *Decision Tree Regressor*. Pengujian yang dilakukan menunjukkan metode KNN *Regressor* mendapatkan nilai RMSE yang paling rendah. Maka metode KNN *Regressor* lebih unggul daripada metode lainnya. Selanjutnya penelitian [15] oleh Nurfauzan dan Fatimah melakukan dengan melakukan prediksi harga saham menggunakan algoritma *K-Nearest Neighbors* (KNN). Hasil dari penelitian tersebut mendapatkan akurasi sebesar 98.8% dengan nilai RMSE 169. Hasil tersebut merupakan hasil yang akurat untuk memprediksi harga saham menggunakan algoritma KNN. Penelitian [18] oleh Seruni, dkk menggunakan metode *K-Nearest Neighbor* (KNN) *Regressor* untuk prediksi jumlah penduduk Kota Malang. Terdapat dua pengujian yaitu pengujian jangka panjang dan pendek. Hasil pengujian jangka pendek mendapatkan hasil MAPE sebesar 0.02526%. Sedangkan pada pengujian jangka panjang, mendapatkan MAPE 0.13506% untuk prediksi selama 24 bulan ke depan. Kedua hasil tersebut membuktikan bahwa metode KNN *Regressor* dapat bekerja dengan akurat untuk prediksi jumlah penduduk di Malang dalam jangka waktu atau 1-3 tahun kedepan. Dari beberapa penelitian diatas, dapat disimpulkan bahwa metode *K-Nearest Neighbors* (KNN)

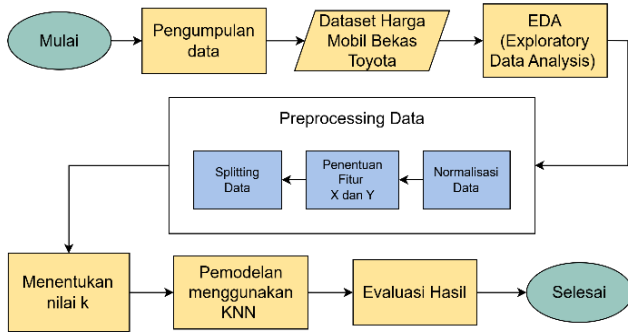
mampu menghasilkan nilai yang lebih akurat digunakan dibandingkan metode lainnya.

Selain itu terdapat beberapa penelitian yang menjelaskan tentang prediksi harga mobil bekas menggunakan model *machine learning*. Penelitian [7] yang dilakukan oleh Kusuma dan Hidayat dengan menerapkan metode *Regresi Linear* dalam memprediksi harga mobil bekas di India. Untuk mendapatkan hasil yang maksimal, penelitian ini melakukan 4 kali pengujian dengan rasio data yang berbeda-beda. Hasil akurasi paling baik yang didapatkan yaitu 71.09% dengan rasio data uji dan data latih 80:20. Akan tetapi hasil tersebut masih cukup rendah sebagai nilai akurasi dalam melakukan prediksi harga mobil bekas di India. Penelitian Putra, dkk merupakan penelitian untuk memprediksi harga mobil menggunakan metode *Random Forest* dan *Decision Tree*. Perbandingan kedua metode tersebut menghasilkan metode *Random Forest* yang lebih unggul nilai akurasinya (72.13%) dibandingkan *Decision Tree* (67.21%). Hasil akurasi dari kedua metode yang digunakan masih terbilang rendah dalam melakukan prediksi harga mobil [12]. Adapun penelitian [19]. Kriswantara dan Sadikin melakukan eksperimen terhadap lima model regresi yaitu *Ridge*, *Decision Tree*, *Random Forest*, *Gradient Boosting*, dan *XGBOOST* untuk melakukan prediksi harga mobil bekas dan peneliti akan mencari model regresi yang tepat dengan akurasi yang baik. Hasil pengujian dengan lima model regresi, *Random Forest* mendapatkan tingkat error MAE dan RMSE yang paling rendah, dengan MAE=1.006 dan RMSE=1,452. Dengan itu, metode *Random Forest* dapat memberikan prediksi regresi pada harga mobil bekas. Dalam konteks prediksi harga mobil bekas, dapat disimpulkan bahwa beberapa metode yang telah dijelaskan belum mampu menghasilkan akurasi yang cukup baik.

Berdasarkan beberapa penelitian yang dijelaskan, metode *K-Nearest Neighbor* (KNN) memiliki kinerja yang baik digunakan. Selain itu metode KNN belum pernah digunakan dalam memprediksi mobil bekas Toyota. Maka, peneliti akan memilih metode KNN *Regressor* untuk memprediksi harga mobil bekas Toyota. Dengan menggunakan KNN *Regressor*, dapat dilakukan identifikasi pola-pola dalam data harga mobil bekas dan memprediksi harga untuk data baru yang belum pernah terlihat sebelumnya. Metode KNN *Regressor* akan mencari hasil prediksi terbaik dengan membandingkan nilai neighbors. Peneliti berharap dapat memberikan wawasan mengenai metode KNN *Regressor* dalam memprediksi harga mobil bekas Toyota. Hasil dari penelitian diharapkan mampu menjadi acuan bagi para pengembang platform jual beli mobil bekas. Dengan demikian, hasil dari penelitian ini dapat memberikan dampak yang cukup luas untuk berbagai pihak yang telah terlibat.

II. METODE

Penelitian ini dilakukan dengan menerapkan salah satu model *machine learning* yaitu algoritma *K-Nearest Neighbors Regressor*. Metode tersebut akan digunakan untuk memprediksi harga mobil bekas Toyota berdasarkan fitur-fitur yang relevan. Adapun alur penelitian yang akan digunakan dalam penelitian ini, yang akan ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Data yang digunakan pada penelitian ini diambil dari Kaggle [20] yang bersifat publik. Dataset tersebut berisi kumpulan informasi tentang mobil bekas Toyota, seperti harga, model mobil, tahun pembuatan, pajak, bahan bakar, jarak tempuh mobil, transmisi, mpg, dan juga cc mobil. Dataset ini berjumlah 409 data, dengan 9 fitur. Dataset ini mencakup mobil Toyota dengan 8 model merek keluaran tahun 2005-2023.

B. Exploratory Data Analysis (EDA)

Pada proses ini pertama dilakukan adalah pengecekan awal dataset. Pengecekan dataset dilakukan untuk menampilkan isi dari dataset, serta informasi mengenai gambaran umum pada dataset.

TABEL I
HASIL TRANSFORMASI DATASET

No	Nama Fitur	Type Data
1	Model	Object
2	Tahun	Int64
3	Harga	Int64
4	Transmisi	Object
5	Kilometer	Int64
6	Bahan Bakar	Object
7	Pajak	Int64
8	Mpg	Float64
9	Cc	Int64
10	Model Encoded	Int64
11	Transmisi Encoded	Int64
12	Bahan Bakar Encoded	Int64

Dalam EDA, diperlukan juga pengecekan informasi mengenai tipe data untuk memahami jenis data. Dataset ini

mempunyai 6 kolom data object dan 3 kolom data string. Selanjutnya melakukan transformasi data. Transformasi ini digunakan untuk mengubah tipe data objek menjadi numerik, karena *machine learning* hanya bisa digunakan pada data numerik [19]. Pada dataset ini menggunakan *Label Encoder* untuk mengubah label kategorik menjadi nilai numerik.

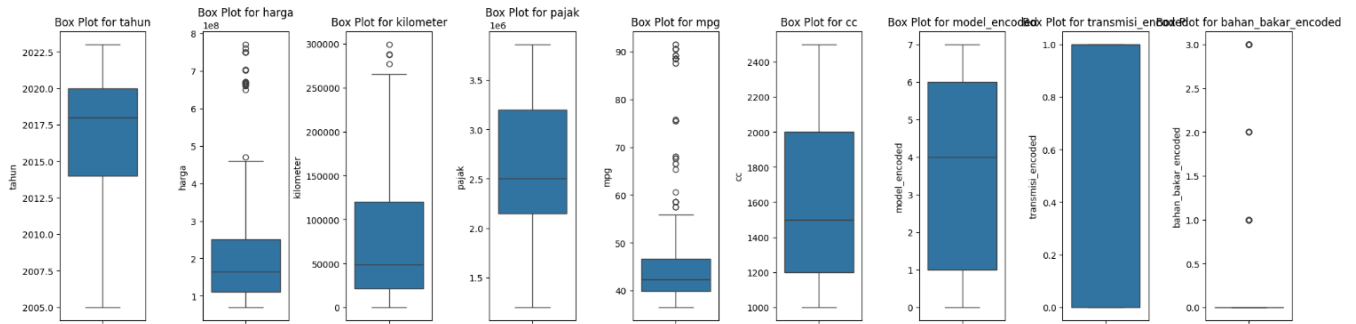
Pada proses transformasi diatas, melakukan penambahan kolom baru dengan menggunakan *Label Encode*. Penyimpanan kolom asli masih diperlukan karena akan digunakan untuk analisis hasil. Hal ini dikarenakan penggunaan data yang di encode hanya dilakukan pada tahap tertentu. Dengan begitu jumlah kolom bertambah menjadi 12 kolom. Seperti pada table I.

Kemudian melakukan pengecekan terhadap statistik deskriptif untuk melihat jumlah data, nilai maksimal dan minimal data, nilai rata-rata, dan standar deviasi untuk setiap kolom numerik. Langkah tersebut berguna untuk mendapatkan ringkasan statistik tentang distribusi data, variasi setiap kolom numerik, dan mengidentifikasi data yang mungkin perlu dilakukan penanganan khusus seperti *outlier*.

Selanjutnya, pengecekan data duplikat dan data yang hilang atau *missing value*. Jika terdapat data duplikat dan *missing value*, maka perlu dilakukan penghapusan data karena dapat mempengaruhi hasil akurasi. Data duplikat dan *missing value* dapat mempengaruhi *machine learning* dalam membuat model pembelajaran dan akurasi yang dihasilkan [19]. Untuk mendapatkan hasil yang lebih akurat, langkah selanjutnya yaitu pengecekan *outlier*. *Outlier* merupakan data yang jauh dari pusat data dan letaknya di luar distribusi normal. Pengecekan *outlier* dapat dilakukan dengan memvisualisasikan fitur dataset dengan menggunakan boxplot [21]. Pada data ini terdapat kolom yang memiliki outlier yang dapat dilihat pada tabel II. Selain itu terdapat boxplot untuk outlier pada gambar 2.

TABEL II
FITUR YANG MEMILIKI OUTLIER

No	Nama Fitur	Outlier	
		Ya	Tidak
1	Model		✓
2	Tahun		✓
3	Harga	✓	
4	Transmisi		✓
5	Kilometer	✓	
6	Bahan Bakar		✓
7	Pajak		✓
8	Mpg		✓
9	Cc	✓	



Gambar 2. Boxplot outlier dan distribusi data

Pada gambar diatas, dapat terlihat distribusi data yang memiliki karakteristik beragam. Panjang garis atau whiskers yang ada di luar kotak menunjukkan persebaran data. Jika whiskers panjang, maka data memiliki persebaran yang luas. Sedangkan whiskers pendek, data lebih terkonsentrasi. Garis tengah yang ada di dalam kotak yaitu median. Batas bawah dan atas menunjukkan Q1 dan Q3 dengan sebaran data 50% di tengah. Jika kotaknya kecil, maka data cenderung memusat, dan jika kotaknya panjang, maka data tersebar merata. Selain itu terdapat titik-titik di luar whiskers yang disebut outlier. Semakin banyak outlier, semakin besar data tidak terdistribusi secara merata. Pada kolom tahun, pajak, cc, model, dan transmisi menunjukkan distribusi data yang merata dan tidak terdapat outlier. Sedangkan kolom harga memiliki outlier yang menonjol. Hal tersebut menunjukkan bahwa terdapat kendaraan yang mempunyai harga yang jauh lebih tinggi dibandingkan mayoritas data. Kolom kilometer juga memiliki outlier yang menandakan terdapat kendaraan dengan jarak tempuh yang ekstrem. Selain itu mpg juga memiliki outlier. Terdapat kendaraan yang mempunyai efisiensi bahan bakar yang jauh lebih tinggi atau lebih rendah dibandingkan yang lainnya. Terdapat empat kolom yang mempunyai outlier yaitu kolom harga, mpg, kilometer, dan transmisi. Ketiga kolom tersebut akan dipertimbangkan untuk dilakukan perhitungan IQR. *Inter-Quartile Range* (IQR) merupakan selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1). Data yang memiliki nilai di atas batas atas atau di bawah batas bawah dianggap sebagai outlier yang kemudian akan di filter atau dihapus dari dataset. Setelah penanganan outlier dilakukan, ukuran dataset yang berubah menjadi 9 kolom 353 baris.

TABEL III
PENANGANAN OUTLIER

Sebelum	Sesudah
12 kolom	12 kolom
409 baris	353 baris

C. Preprocessing Data

Preprocessing data yaitu langkah penting yang perlu dilakukan sebelum melakukan proses pemodelan [22]. Terdapat beberapa tahap :

1) Normalisasi Data

Normalisasi data juga perlu dilakukan karena terdapat kolom yang mempunyai skala nilai yang besar. Data numerik seperti jarak tempuh, harga, dan pajak dilakukan penskalaan untuk memastikan skala yang sama.

2) Penentuan Fitur X dan Y

Selanjutnya data akan dibagi menjadi dua, yaitu variabel independen (X) dan dependen (Y). Variabel X berisi data input yang akan digunakan model untuk melakukan prediksi. Sedangkan variabel Y berisi nilai yang ingin diprediksi dengan model yang telah dipilih.

3) Splitting Data

Pada tahap ini, data akan dibagi dalam dua bagian sebagai pembelajaran kinerja model dengan perbandingan 80:20, 80% untuk data *train* dan 20% untuk data *test*.

D. K-Nearest Neighbor Regressor

K-Nearest Neighbor (KNN) merupakan salah satu metode yang digunakan untuk melakukan klasifikasi objek berdasarkan data pelatihan dengan jarak terdekat dengan objek tersebut [18]. Meskipun KNN umumnya digunakan untuk klasifikasi, tetapi metode ini juga dapat juga digunakan untuk regresi, yang sering disebut sebagai KNN Regression. K-Nearest Neighbor Regressor merupakan hasil regresi yang didapatkan dari rata-rata k tetangga (*neighbor*) terdekat [18].

$$y(t + 1) = \frac{1}{k} \sum_i^k = 1y_i \tag{1}$$

Keterangan :

- $y(t+1)$ = data yang diprediksi
- k = jumlah tetangga
- y_i = data ke-i

Algoritma *K-Nearest Neighbor (KNN) Regressor* diterapkan untuk melatih model berdasarkan data latih yang telah diolah. KNN bekerja dengan mencari sejumlah tetangga terdekat dari data baru untuk memprediksi harga berdasarkan data yang mirip. Penelitian ini jumlah tetangga ($n_neighbors$) yang digunakan yaitu 1, jadi prediksi dilakukan berdasarkan satu tetangga terdekat.

E. Evaluasi Model

Pada evaluasi model, performa akan dievaluasi menggunakan beberapa metrik pengukuran.

1) Mean Squared Error (MSE)

MSE adalah metrik utama yang digunakan dalam regresi untuk mengukur sejauh mana nilai prediksi model berbeda dari nilai aktual. MSE akan memberikan bobot lebih besar pada kesalahan yang lebih besar, karena menghitung dari kesalahan [23].

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

Keterangan :

- n = jumlah data
- y_i = nilai aktual
- \hat{y}_i = nilai prediksi

2) Mean Absolute Error (MAE)

MAE digunakan untuk regresi karena dapat memberikan rata-rata kesalahan dalam satuan variabel. MAE merupakan rata-rata selisih nilai mutlak dengan nilai prediksi. MAE digunakan untuk mengukur tingkat keakuratan model dalam melakukan prediksi [24].

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

Keterangan :

- n = jumlah data
- y_t = nilai aktual
- \hat{y}_t = nilai prediksi

3) R-squared (R^2)

R^2 merupakan *R-Squared* merupakan indikator yang menunjukkan seberapa besar kontribusi variabel independen terhadap variabel dependen. Nilai ini berkisar antara 0 hingga 1, di mana semakin mendekati 1 berarti semakin kuat pengaruh variabel-variabel independen secara kolektif dalam menjelaskan variasi dari variabel dependen [24].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_t - \hat{y}_t)^2}{\sum_{i=1}^n (y_t - \bar{y})^2} \quad (4)$$

Keterangan :

- y_t = nilai aktual
- \hat{y}_t = nilai prediksi
- \bar{y} = rata-rata dari nilai aktual
- n = banyak data

4) Root Mean Squared Error (RMSE)

RMSE merupakan metode dengan menghitung selisih dari nilai aktual dan nilai prediksi. Nilai ini hasil dari akar kuadrat dari perhitungan MSE [25]. Keakuratan pada estimasi sebuah nilai akan dibuktikan dengan hasil RMSE yang kecil. Penelitian ini menggunakan parameter RMSE untuk menilai tingkat akurasi suatu model [15].

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}} \quad (5)$$

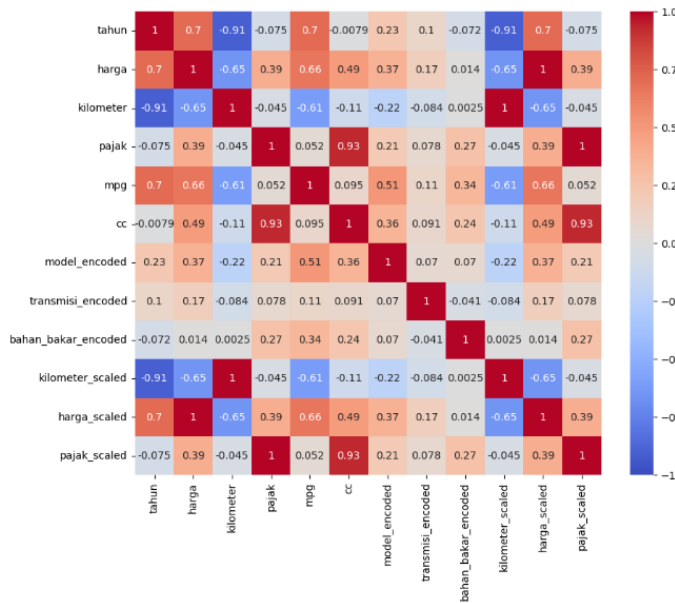
Keterangan :

- Y_t = nilai aktual
- \hat{Y}_t = nilai prediksi
- n = banyak data

III. HASIL DAN PEMBAHASAN

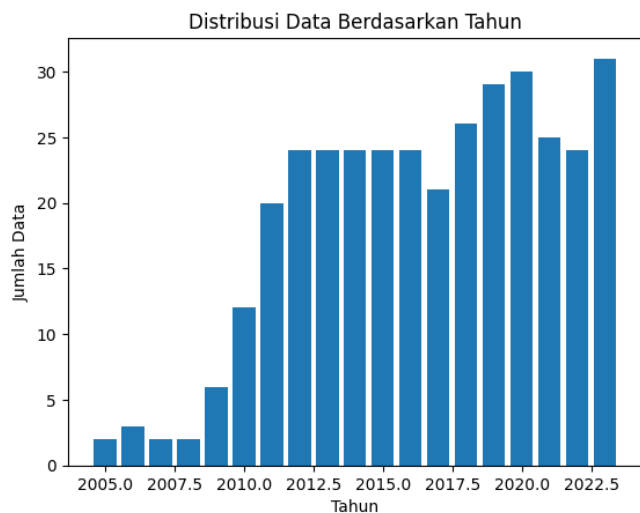
Penelitian ini dilakukan untuk membangun model *K-Nearest Neighbor (KNN) Regressor* dalam melakukan prediksi harga mobil bekas Toyota. Penelitian ini menggunakan dataset dari situs Kaggle dengan judul Mobil Toyota. Dalam dataset hanya mencakup keluaran mobil Toyota pada tahun 2005-2023 dengan 8 merek mobil Toyota. Setelah mendapatkan dataset, peneliti akan melakukan proses EDA. Pada tahap ini, peneliti akan melakukan pengecekan awal pada dataset. Tidak ditemukan data duplikat dan data kosong (*missing value*). Selain itu dilakukan transformasi data dan pengecekan outlier. Transformasi dilakukan pada 3 kolom yang bersifat object, yaitu model, transmisi, dan bahan bakar. Ditemukan juga outlier pada 4 kolom, yaitu harga, mpg, bahan bakar, dan kilometer. Dan sudah dilakukan proses penanganan outlier. Selanjutnya dataset akan di proses melalui beberapa tahap *Preprocessing*, seperti normalisasi data, pembagian variable, dan *splitting data*.

Dalam proses normalisasi data, penskalaan dilakukan terhadap kolom yang memiliki rentang nilai yang besar, seperti kolom kilometer, harga, dan pajak. Tahap ini dilakukan untuk mengurangi perbedaan skala dan agar nilai kolom-kolom tersebut lebih mudah untuk dianalisis. Selanjutnya data yang telah diolah akan dibagi menjadi dua bagian, variabel independen (X) dan dependen (Y). Variabel X akan berisi data input yang digunakan model dalam melakukan prediksi. Sedangkan variabel Y berisi nilai yang ingin diprediksi dengan model yang telah dipilih. Pada tahap ini, pemilihan variabel X dan Y dilakukan dengan menganalisis hasil korelasi. Hasil korelasi dapat dilihat pada gambar 3.

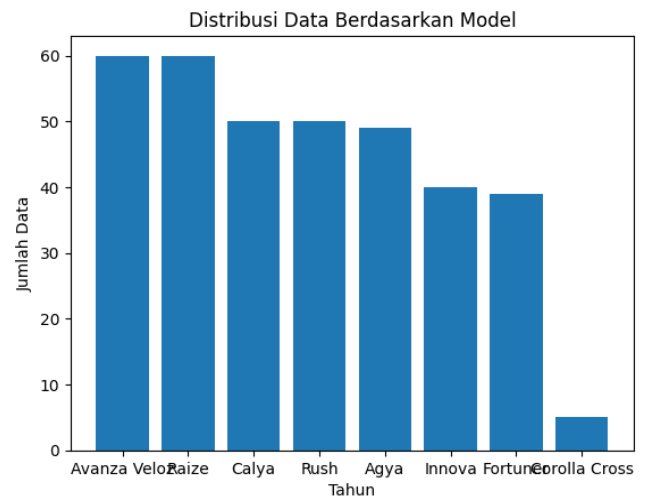


Gambar 3. Hasil Korelasi

Dari gambar 2, hasil korelasi tersebut diperoleh fitur relevan yang dapat mempengaruhi harga mobil bekas, yakni Model, Tahun, Kilometer, Pajak, Mpg, dan Cc. Fitur relevan tersebut dipilih dengan batasan 0.3 sampai 1 dan -0.3 sampai -1. Fitur transmisi dan bahan bakar menunjukkan korelasi yang sangat rendah. Jika fitur tersebut digunakan, performa dari model akan mengalami penurunan. Fitur yang paling berpengaruh yaitu pada kolom tahun dengan nilai korelasi 0.7. Terdapat distribusi jumlah mobil dan model mobil berdasarkan tahun pembuatan pada gambar 4 dan gambar 5.



Gambar 4. Distribusi data berdasarkan tahun



Gambar 5. Distribusi data berdasarkan model

Pada gambar 4 menjelaskan bahwa data hanya menggunakan kendaraan dengan tahun produksi mulai dari 2005 sampai 2023. Jumlah kendaraan meningkat secara signifikan mulai dari 2009 sampai tahun-tahun berikutnya. Hal ini menunjukkan bahwa data tersebut berfokus pada kendaraan yang tahun pembuatannya lebih baru. Sedangkan pada gambar 5 menunjukkan bahwa model mobil yang digunakan pada penelitian ini hanya 8 jenis merek mobil.

Setelah diperoleh fitur yang relevan, maka dapat menentukan variabel X dan Y. Variabel X berisi kolom yang relevan dan Y berisi kolom harga. Kemudian data akan dibagi tiga bagian dengan perbandingan yang berbeda-beda untuk memastikan keakuratan model evaluasi. Perbandingan yang akan digunakan dapat dilihat pada tabel IV.

TABEL IV
PERBANDINGAN YANG DIGUNAKAN

Perbandingan	Train	Test
90:10	90	10
80:20	80	20
70:30	70	30
10 Fold Validasi	90	10
5 Fold Validasi	80	20

Data yang telah dibagi selanjutnya akan diterapkan ke dalam model *K-Nearest Neighbor* (KNN) *Regressor* dengan melihat hasil yang terbaik dari pembagian data yang digunakan yaitu *splitting* data dan *cross-validation* atau *KFold*. Pengujian ini menggunakan 4 metrik seperti MAE, MSE, RMSE, dan R2-score. Hasil *splitting* dapat dilihat pada tabel V.

TABEL V
HASIL EVALUASI SPLITTING DATA

Perbandingan	MAE	MSE	RMSE	R2-Score
90:10	5.81757	75.80733	8.70677	0.98569
80:20	5.91135	98.98308	9.94902	0.98640
70:30	7.45541	173.25014	13.16245	0.97347
10 Fold Validasi	11.14	608.13	24.66	0.92
5 Fold Validasi	11.24	613.17	24.76	0.92

Hasil di atas menunjukkan nilai MAE, MSE, dan RMSE dengan perbandingan 90:10 lebih rendah dibandingkan dengan perbandingan lainnya. Ketiga metrik error tersebut menunjukkan besar perbedaan rata-rata antara nilai prediksi dan nilai aktual. Meskipun nilai R2-Score pada perbandingan 80:20 lebih besar, tetapi perbandingan 90:10 merupakan pilihan terbaik karena dalam prediksi harga nilai MSE, MAE, dan RMSE yang kecil dapat menjamin hasil prediksi yang lebih akurat. Jadi penelitian ini akan menggunakan perbandingan 90:10 dalam memprediksi harga mobil bekas. Data yang telah dibagi selanjutnya akan diterapkan dalam model *K-Nearest Neighbor* (KNN) *Regressor* dengan menentukan nilai k. Pengujian nilai k dilakukan dengan cara menguji secara manual pada berbagai nilai k untuk mengetahui hasil yang optimal. Kemudian data akan dievaluasi kembali menggunakan parameter MAE, MSE, RMSE, dan R2-Score. Hasil pengujian dapat dilihat pada tabel VI.

TABEL VI
HASIL PEMODELAN KNN REGRESSOR

Nilai k=	MAE	MSE	RMSE	R2-Score
1	3.31686	26.43640	5.14163	0.99501
2	4.24472	38.89976	6.23696	0.99265
3	5.14533	49.09838	7.00702	0.99073
4	4.69936	40.83867	6.39051	0.99229
5	5.81757	75.80733	8.70674	0.98569
6	7.25341	145.26456	12.05257	0.97258
7	8.67707	244.36029	15.63202	0.95387
8	9.34637	316.32005	17.78538	0.94029
9	9.42041	361.71192	19.01872	0.93173
10	9.42041	405.98281	20.14901	0.92337

Berdasarkan hasil pengujian nilai k=1 sampai k=10, didapatkan hasil yang paling baik yaitu nilai k=1. Pengujian dengan nilai k=1 mendapatkan hasil MAE, MSE, dan RMSE yang paling kecil, sedangkan nilai R2-Score mendapatkan nilai paling tinggi. Hasil nilai MAE sebesar 3.31686. Dalam konteks harga mobil bekas, nilai tersebut merupakan rata-rata selisih absolut antara harga yang diprediksi dan harga sebenarnya yaitu sekitar 3 juta rupiah. Nilai MSE 26.43640 menunjukkan rata-rata kuadrat selisih antara harga prediksi dan harga aktual sebesar 26.4 juta rupiah. Harga mobil bekas ini berada dalam skala yang besar seperti ratusan juta rupiah. Nilai RMSE sebesar 5.14163 menunjukkan kesalahan rata-rata dalam satuan yang sama, atau sekitar 5.1 juta rupiah. Terakhir yaitu R2-Score dengan nilai 0.99501. Hal ini membuktikan bahwa model KNN *Regressor* dapat menjelaskan 99.50% variabilitas dalam harga mobil. Nilai tersebut merupakan nilai yang akurat dan baik digunakan untuk memprediksi harga mobil bekas dengan menggunakan model *K-Nearest Neighbor* (KNN) *Regressor*.

TABEL VII
PERBANDINGAN DENGAN PENELITIAN SEBELUMNYA

Spesifikasi	Linear Regression	Random Forest dan Decision Tree	KNN Regressor
Sumber Dataset	Kaggle	Kaggle	Kaggle
Jumlah data	6.019 baris 14 kolom	205 baris 9 kolom	409 baris 9 kolom
Akurasi	71.09%	RF = 72.13% DT = 67.21%	99.50%

IV. KESIMPULAN

Penelitian ini mengimplementasikan metode *K-Nearest Neighbor* (KNN) *Regressor* dalam memprediksi harga mobil bekas Toyota dengan dataset yang diambil dari Kaggle. Penelitian ini berhasil mengolah melalui beberapa tahap seperti pengumpulan data, EDA, dan *preprocessing*. Fitur yang digunakan juga akan dipertimbangkan menggunakan korelasi dengan batasan 0.3 sampai 1 dan -0.3 sampai -1 agar dapat meningkatkan hasil evaluasi. Fitur yang digunakan yaitu Model, Tahun, Kilometer, Pajak, Mpg, dan Cc. Setelah mencari perbandingan pembagian data, 90:10 merupakan hasil perbandingan terbaik dibandingkan menggunakan 80:20, 70:30 dan pengujian K Fold dengan 10 fold dan 5 fold. Selanjutnya dalam penentuan nilai k, hasil evaluasi menunjukkan bahwa nilai k=1 dapat memberikan performa paling baik dengan menggunakan metode *K-Nearest Neighbor* (KNN) *Regressor*. Hasil dari k=1 diperoleh nilai MAE 3.31686, MSE 26.43640, RMSE 5.14163, dan R2-Score sebesar 0.99501. Nilai R2-Score menunjukkan bahwa tingkat akurasi yang dihasilkan untuk memprediksi harga mobil bekas Toyota sebesar 99,46%. Hal ini membuktikan bahwa KNN *Regressor* merupakan metode yang efektif dalam memprediksi harga mobil bekas Toyota. Selain itu metode KNN memberikan nilai akurasi paling tinggi dibandingkan dengan peneliti sebelumnya. Pada penelitian ini model yang digunakan hanya 8 merek mobil Toyota, seperti pada gambar 5. Dan tahun pembuatan mobil hanya dari 2005-2023. Hal ini memungkinkan terjadi penurunan atau kenaikan pada performanya jika menggunakan dataset yang berbeda. Penelitian ini diharapkan mampu menjadi acuan bagi para pengembang platform jual beli mobil bekas. Dengan demikian, hasil dari penelitian ini dapat memberikan dampak yang cukup luas untuk berbagai pihak.

DAFTAR PUSTAKA

- [1] P. A. Widyasmoro and S. Suryoko, "Pengaruh Harga, Kualitas Produk dan Promosi Terhadap Keputusan Pembelian Mobil Merek Toyota Sienta di PT. Nasmoco Gombel Semarang," *J. Ilmu Adm. Bisnis*, vol. 9, no. 4, pp. 431-438, 2020, doi: 10.14710/jiab.2020.28713.
- [2] Infootomotif, "Info Otomotif: Manfaat Mobil untuk Manusia." [Online]. Available: <https://kumparan.com/info-otomotif/manfaat-mobil-untuk-aktivitas-manusia-1wqysMbGRz0/1>
- [3] P. N. M. Yoedo, "Pasar Mobil Domestik," 2023, [Online]. Available: <https://otomotif.bisnis.com/read/20231210/275/1722540/pasar->

- mobil-domestik-lesu-begini-proyeksi-penjualan-toyota-hingga-akhir-2023
- [4] A. Hadian, "Pengaruh Harga Produk Dan Mencari Variasi Terhadap Terhadap Perpindahan Merek Mobil Mpv Di Kota Medan," *J. Penelit. Pendidik. Sos. Hum.*, vol. 3, no. 1, pp. 346–354, 2018, doi: 10.32696/jp2sh.v3i1.98.
- [5] A. I. Lestari, Verawaty, Setya Ega Susanto, Baharuddin, and Budhi Krisnanto, "Analisis Faktor Kebudayaan, Sosial, Pribadi Dan Psikologi Terhadap Keputusan Konsumen Membeli Mobil Bekas Pada Dealer Di Kota Makassar," *J. Sains Manaj. Nitro*, vol. 1, no. 2, pp. 182–192, 2022, doi: 10.56858/jsmn.v1i2.98.
- [6] Merdeka.com/OTO, "Menentukan Harga Jual Mobil Bekas agar Tidak Mengalami Kerugian," p. 222330, [Online]. Available: <https://www.merdeka.com/otomotif/begini-caranya-menentukan-harga-jual-mobil-bekas-agar-tidak-mengalami-kerugian-222330-mvk.html>
- [7] 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/jimik.v5i2.629.
- [8] Z. Sun, "Research on factors affecting second-hand car market prices," *Theor. Nat. Sci.*, vol. 36, no. 1, pp. 128–135, 2024, doi: 10.54254/2753-8818/36/20240532.
- [9] E. S. Eriana and D. A. Zein, "Artificial Intelligence," *Angew. Chemie Int. Ed.*, vol. 6(11), p. 1, 2023.
- [10] E. Hasibuan *et al.*, "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.
- [11] A. Amalia, M. Radhi, S. H. Sinurat, D. R. H. Sitompul, and E. Indra, "Prediksi Harga Mobil Menggunakan Algoritma Regresi 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.
- [12] P. H. Putra, A. Azanuddin, B. Purba, and Y. A. Dalimunthe, "Random forest and decision tree algorithms for car price prediction," *J. Mat. Dan Ilmu Pengetah. Alam LLDikti Wil. 1*, vol. 4, no. 1, pp. 81–89, 2023, doi: 10.54076/jumpa.v3i2.305.
- [13] G. Budiprasetyo, M. Hani'ah, and D. Z. Aflah, "Prediksi Harga Saham Syariah Menggunakan Algoritma Long Short-Term Memory (LSTM)," *J. Nas. Teknol. dan Sist. Inf.*, vol. 8, no. 3, pp. 164–172, 2023, doi: 10.25077/teknosi.v8i3.2022.164-172.
- [14] R. Rahmadini, Enjel Erika LorencisLubis, Aji Priansyah, Yolanda R.W.N, and Tuti Meutia, "Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor," *J. Mhs. Akunt. Samudra*, vol. 4, no. 4, pp. 223–235, 2023, doi: 10.33059/jmas.v4i4.7074.
- [15] D. Nurfauzan and T. Fatimah, "Implementasi Algoritma K-Nearest Neighbors Regression Dalam Memprediksi Harga Saham," *Semin. Nas. Mhs.*, no. September, pp. 576–584, 2022, [Online]. Available: <http://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/391%0Ahttps://senafiti.budiluhur.ac.id/index.php/senafiti/article/download/391/67>
- [16] Mukhlisin, M. Imrona, and D. T. Murdiansyah, "Prediksi Harga Beras Premium dengan Metode Algoritma K-Nearest Neighbor," *e-Proceeding Eng.*, vol. 7, no. 1, pp. 2714–2724, 2019.
- [17] U. M. Area, "Pengertian KNN." 2023. [Online]. Available: <https://lp2m.uma.ac.id/2023/02/16/algoritma-k-nearest-neighbors-knn-pengertian-dan-penerapan/>
- [18] D. S. Seruni, M. T. Furqon, and R. C. Wihandika, "Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode K-Nearest Neighbor Regression," *Sist. Prediksi Pertumbuhan Jumlah Pendud. Kota Malang menggunakan Metod. K-Nearest Neighbor Regres.*, vol. 4, no. 4, pp. 1075–1082, 2020.
- [19] B. Kriswantara and R. Sadikin, "Used Car Price Prediction with Random Forest Regressor Model," *J. Inf. Syst. Informatics Comput. Issue Period*, vol. 6, no. 1, pp. 40–49, 2022, doi: 10.52362/jisicom.v6i1.752.
- [20] M. Raynold, "Dataset Harga Mobil Toyota." 2022. [Online]. Available: <https://kaggle.com/datasets/muhammadraynold/datasetmobil>
- [21] A. M. M. Fattah, A. Voutama, N. Heryana, and N. Sulistiyowati, "Pengembangan Model Machine Learning Regresi sebagai Web Service untuk Prediksi Harga Pembelian Mobil dengan Metode CRISP-DM," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1669, 2022, doi: 10.30865/jurikom.v9i5.5021.
- [22] J. Setyanto and T. B. Sasongko, "Sentiment Analysis of Sirekap Application Users Using the Support Vector Machine Algorithm," *J. Appl. Informatics Comput.*, vol. 8, no. 1, pp. 71–76, 2024, doi: 10.30871/jaic.v8i1.7772.
- [23] H. Nuha, "Mean Squared Error (MSE) dan Penggunaannya," *Papers.Ssrn.Com*, vol. 52, pp. 1–1, 2023, [Online]. Available: <https://ssrn.com/abstract=4420880>
- [24] A. T. Nurani, A. Setiawan, and B. Susanto, "Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma," *J. Sains dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, 2023, doi: 10.24246/juses.v6i1p34-43.
- [25] M. A. Saputra, U. Hayati, T. Informatika, M. Informatika, R. Linier, and D. Mining, "Estimasi Harga Mobil Bekas Toyota Yaris menggunakan Algoritma Regresi Linier," vol. 8, no. 2, pp. 1696–1701, 2024.