LINIER
LITERATUR INFORMATIKA & KOMPUTER
e-ISSN: 3063-2218
DOI : 10.33096/linier.v2i4.3346
Literatur Informatika & Komputer
Vol 2, No 4 (2025), pp 606-615

Analisis Faktor Penentu Harga Mobil Bekas Menggunakan Model Random Forest Regressor serta Perbandingan Linear

Sabrina Khoirunnisa[a], Nining Rahaninsih[b], Irfan Ali[c], Willy Prihartono[d]
STMIK IKMI, Cirebon, Indonesia
[a]sabrinanisakhoirunnisa@gmail.com; [b]niningr157@yahoo.co.id; [c]irfanaali0.0@gmail.com; [d]willyprihartono@gmail.com

Received: 21-11-2025 | Revised: 30-11-2025 | Accepted: 05-12-2025 | Published: 29-12-2025

Abstrak
Penentuan harga mobil bekas merupakan proses yang kompleks karena dipengaruhi oleh berbagai faktor seperti jarak tempuh, usia kendaraan, merek, jenis bahan bakar, dan riwayat kerusakan. Metode konvensional sering kali menghasilkan penilaian yang subjektif dan kurang akurat, sehingga diperlukan pendekatan berbasis data untuk meningkatkan objektivitas dan konsistensi estimasi harga. Penelitian ini bertujuan untuk membangun model prediksi harga mobil bekas menggunakan algoritma Random Forest Regressor serta membandingkan performanya dengan Multiple Linear Regression sebagai baseline. Dataset yang digunakan berasal dari platform Kaggle dan mencakup 4.009 data kendaraan yang telah melalui proses data cleaning, rekayasa fitur, dan penghapusan outlier. Metode penelitian meliputi preprocessing data, pelatihan model, hyperparameter tuning, serta evaluasi menggunakan metrik R^2, MAE, MSE, dan RMSE. Hipotesis penelitian menyatakan bahwa Random Forest memiliki performa prediktif yang lebih baik dibandingkan model linier serta mampu mengidentifikasi fitur yang paling berpengaruh terhadap harga kendaraan. Hasil eksperimen menunjukkan bahwa Random Forest $R^2 = 0.6827$, lebih tinggi dibandingkan Multiple Linear Regression dengan $R^2 = 0.5673$. Analisis feature importance mengungkapkan bahwa mileage dan usia kendaraan merupakan faktor dominan dalam pembentukan harga. Dengan demikian, penelitian ini menyimpulkan bahwa Random Forest merupakan pendekatan yang lebih akurat dan stabil untuk prediksi harga mobil bekas serta berpotensi diimplementasikan dalam sistem valuasi otomotif berbasis data.

Kata kunci: prediksi harga mobil, machine learning, random forest, regresi linier, feature importance

Pendahuluan

Perkembangan teknologi informasi dan komunikasi dalam dekade terakhir telah mendorong transformasi signifikan di berbagai sektor, termasuk industri otomotif. Kemampuan untuk memproses data dalam skala besar menjadi elemen penting dalam mendukung pengambilan keputusan berbasis informasi. Dalam konteks ini, ilmu informatika berperan besar dalam menyediakan solusi teknologi yang mampu mengatasi permasalahan nyata, salah satunya melalui pemanfaatan *machine learning* dalam proses estimasi harga kendaraan. Berbagai studi menunjukkan bahwa algoritma prediktif memiliki potensi besar dalam memberikan estimasi harga mobil bekas yang lebih akurat dan objektif dibanding pendekatan manual [1], [2].

Namun demikian, proses penentuan harga mobil bekas masih menghadapi tantangan, khususnya terkait kompleksitas faktor yang memengaruhi nilai jual kendaraan. Harga mobil bekas dipengaruhi oleh usia kendaraan, jarak tempuh, merek, kondisi fisik, jenis bahan bakar, hingga riwayat kecelakaan. Sistem penilaian konvensional sering kali belum mempertimbangkan hubungan non-linier antar variabel tersebut, sehingga berpotensi menghasilkan estimasi harga yang bias dan tidak akurat. Pendekatan komputasional berbasis data diperlukan agar proses penilaian dapat dilakukan secara lebih objektif dan presisi [3], [4].

Penelitian terdahulu telah mengaplikasikan berbagai algoritma *machine learning* untuk memprediksi harga mobil bekas. [5] menunjukkan bahwa metode ensemble seperti Random Forest memberikan performa lebih baik dibanding model linier. [6] menggunakan regresi linier sebagai baseline yang interpretatif namun terbatas dalam menangkap interaksi kompleks antar fitur. Sementara itu, [7] menemukan bahwa model *boosting* seperti XGBoost dan LightGBM menawarkan akurasi yang lebih tinggi melalui pemrosesan fitur yang lebih komprehensif. Meski demikian, sebagian studi masih minim dalam aspek *preprocessing*, rekayasa fitur, penanganan outlier, serta penyesuaian model terhadap karakteristik data lokal.

Di tengah meningkatnya permintaan pasar mobil bekas, baik secara global maupun domestik, konsumen dan pelaku industri membutuhkan sistem penilaian harga yang akurat dan konsisten. Sifat subjektif dalam proses penaksiran harga sering kali menjadikan keputusan kurang tepat. Mengingat hubungan antar variabel seperti

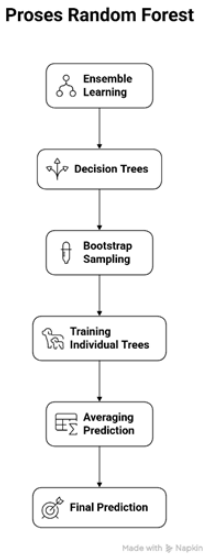
car age, *mileage*, *brand*, dan *fuel type* bersifat kompleks serta tidak selalu linier, maka algoritma berbasis *ensemble learning* menjadi alternatif yang menjanjikan.

Berdasarkan kebutuhan tersebut, penelitian ini mengembangkan model prediksi harga mobil bekas menggunakan algoritma Random Forest Regressor dan membandingkannya dengan Multiple Linear Regression sebagai baseline. Penelitian dilakukan menggunakan dataset berisi lebih dari 4.000 data historis dari platform Kaggle, melalui tahapan *data cleaning*, *exploratory data analysis*, *feature engineering*, pelatihan model, serta evaluasi menggunakan metrik R^2, MAE, MSE, dan RMSE. Teknik *hyperparameter tuning* melalui GridSearchCV digunakan untuk mengoptimalkan kinerja model.

Apabila model yang dikembangkan mampu mencapai tingkat akurasi yang tinggi, maka hasil penelitian ini dapat menjadi dasar bagi pengembangan sistem valuasi harga mobil bekas yang otomatis dan berbasis data. Selain itu, penelitian ini diharapkan dapat berkontribusi pada pengembangan teknologi informasi di sektor otomotif dan membuka peluang penelitian lanjutan yang berfokus pada prediksi harga aset lain seperti properti dan barang.

Metode

Gambar 1 menunjukkan arsitektur umum alur kerja machine learning yang menjadi acuan dalam penelitian ini, mencakup tahapan akuisisi data, pra-pemrosesan, EDA, pemodelan, evaluasi, dan interpretasi fitur. Pendekatan berlapis ini sesuai dengan kerangka penelitian machine learning modern yang digunakan dalam studi-studi sebelumnya [3], [4], [8].



Gambar 1 Alur kerja Random Forest

A. Dataset dan Pra-Pemrosesan

1. Sumber Dataset
   Data diperoleh dari dataset publik *Used Car Prices* di platform Kaggle, berisi 4.009 entri mobil bekas dengan fitur seperti price, mileage, brand, model_year, fuel_type, dan accident. Dataset ini dipilih karena lengkap, representatif, dan telah digunakan pada penelitian serupa sehingga valid untuk studi prediksi harga.

2. Tahapan Pra-Pemrosesan Data
   Pra-pemrosesan dilakukan agar dataset siap digunakan model regresi. Langkah-langkahnya:

   a. Data Cleaning
      Atribut price dan mileage awalnya berupa string dengan simbol dolar dan satuan jarak. Karakter non-numerik dibersihkan menjadi variabel numerik baru: price_clean dan mileage_clean.

b. Imputasi Nilai Hilang
Nilai kosong pada fuel_type dan accident diisi menggunakan mode imputation, yaitu kategori paling sering muncul.

c. Rekayasa Fitur (Feature Engineering)
Dibuat fitur baru:

$$Usia\_Mobil = 2025 - model\_year$$

d. Outlier Filtering (Sangat Penting)
Terdapat 37 data hyper-car (harga > $275.000). Setelah dihapus, R^2 meningkat dari 0.11 menjadi 0.6827.
Outlier dihapus untuk menjaga stabilitas model.

B. Algoritma dan Formulasi Model

Penelitian ini menggunakan dua algoritma regresi, yaitu *Multiple Linear Regression* (MLR) dan *Random Forest Regressor* (RFR). MLR dipilih sebagai model *baseline* karena sifatnya yang transparan dan umum digunakan dalam prediksi harga aset [6], [9]. Sementara itu, RFR digunakan sebagai model utama karena mampu menangani hubungan non-linier, interaksi antar fitur, serta lebih stabil terhadap *noise* dan *overfitting* dibanding model linier tradisional [8], [10], [11].

1. Multiple Linear Regression
MLR mencoba memodelkan hubungan linear antara variabel prediktor dan harga mobil bekas. Bentuk umum persamaan regresi linier diberikan pada Persamaan (1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (1)$$

dengan:

a. (Y) = harga mobil bekas (*price_clean*),

b. $(X_1, X_2, \ldots X_n)$ = fitur prediktor seperti *Car_Age* dan *mileage_clean*,

c. (β_0) = intercept,

d. (β_n) = koefisien regresi.

MLR dipakai untuk mengukur sejauh mana hubungan linear dapat menjelaskan variasi harga kendaraan, serta menjadi pembanding akurasi bagi model ensemble.

2. Random Forest Regressor
RFR merupakan algoritma *ensemble learning* berbasis *bagging* yang menggabungkan banyak *decision tree* untuk menghasilkan prediksi yang stabil dan akurat. Prediksi akhir merupakan rata-rata dari output seluruh pohon pada agregasi regresi sebagaimana ditunjukkan pada Persamaan (2):

$$\hat{Y} = \frac{1}{T}\sum_{t=1}^{T} f_t(X) \quad (2)$$

dengan:

a. (T) = jumlah pohon keputusan,
b. $\left(f_t(X)\right)$ = prediksi dari pohon ke-t,
c. $\left(\hat{Y}\right)$ = prediksi harga kendaraan.

RFR dipilih karena kemampuannya mengatasi non-linearitas, meminimalkan overfitting, serta menyediakan analisis *feature importance* untuk mengidentifikasi fitur paling berpengaruh.

C. Evaluasi Model

Model dievaluasi menggunakan empat metrik:

1. MAE (*Mean Absolute Error*), mengukur rata-rata *error* absolut dari nilai prediksi dan aktual [12].
2. MSE (*Mean Squared Error*), mengukur rata-rata kuadrat dari *error*. Lebih sensitif terhadap error besar [13].

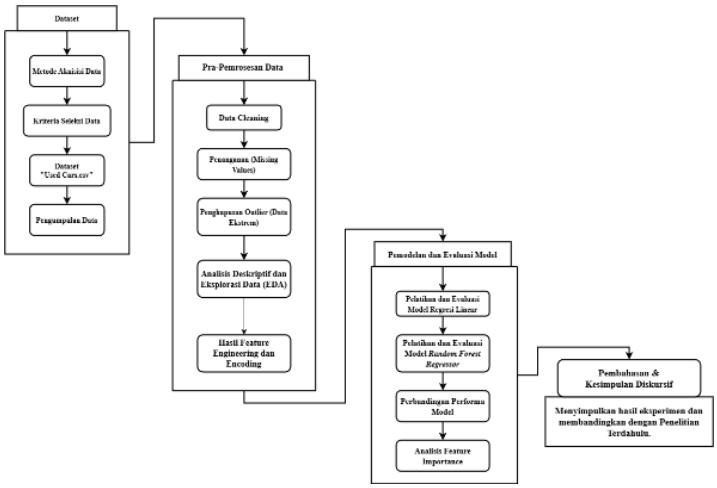
3. RMSE (*Root Mean Squared Error*), akar dari MSE, dengan satuan yang sama seperti data asli sehingga lebih mudah diinterpretasikan [14].
4. R^2 (*Coefficient of Determination*), mengukur kemampuan model menjelaskan variasi data. Nilai mendekekati 1 menunjukkan model semakin baik [15].

Keempat metrik memberikan gambaran menyeluruh mengenai akurasi prediksi dan kemampuan model menjelaskan variasi harga mobil bekas .

Perancangan

Bagian ini menjelaskan alur kerja penelitian, analisis proses, serta rancangan arsitektur sistem prediksi harga mobil bekas. Perancangan dilakukan secara terstruktur mulai dari pengumpulan data, pra-pemrosesan, eksplorasi data, pemodelan, hingga evaluasi performa model.

A. Alur Penelitian (Flowchart Penelitian)
Gambar 3.2 berikut menunjukkan alur tahapan penelitian, mulai dari akuisisi data hingga evaluasi model. Flowchart ini disesuaikan dengan pipeline standar machine learning dan juga digunakan sebagai landasan implementasi pada skripsi.



Gambar 2 .Flowchart Tahapan Metode Penelitian

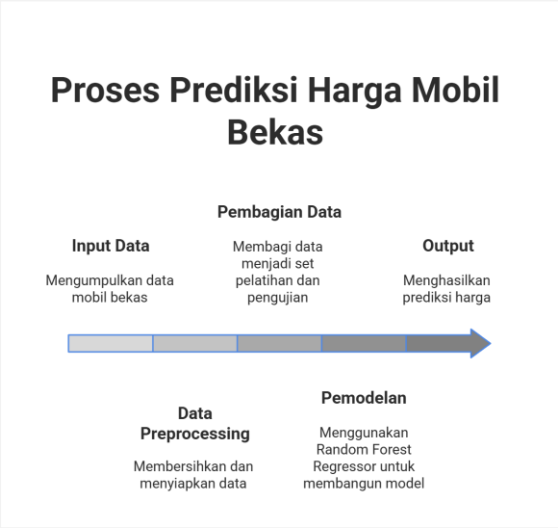
Penjelasan Alur Flowchart:

1. Pengumpulan Data
   Data diambil dari dataset publik Kaggle berisi 4.009 entri mobil bekas. Dataset mencakup variabel harga, mileage, model_year, brand, fuel_type, accident, transmission, dan lainnya.
2. Pra-pemrosesan Data
   Tahap ini mencakup:
   a. *Data cleaning*: menghapus simbol $ dan satuan “mi” pada mileage
   b. *Imputasi nilai hilang*: mode imputation untuk fuel_type dan accident
   c. *Outlier filtering*: menghapus 37 data hyper-car → meningkatkan R^2 dari 0.11 → 0.68
   d. *Feature engineering*: membuat variabel Usia_Mobil = 2025 - model_year
3. Eksplorasi Data (EDA)
   Dilakukan analisis distribusi harga, mileage, usia kendaraan, serta korelasi antar variabel untuk mengetahui pola penting sebelum melakukan pemodelan.
4. Pemodelan Machine Learning
   Dua model dilatih:
   a. Multiple Linear Regression (baseline)
   b. Random Forest Regressor (model utama)
      Model dilatih menggunakan 80% data latih dan diuji pada 20% data uji.

5. Evaluasi Model
   Kinerja kedua model dibandingkan menggunakan MAE, MSE, RMSE, dan R^2.
   Model terbaik dipilih berdasarkan stabilitas, akurasi, serta nilai kesalahan terendah.
6. Interpretasi Fitur (Feature Importance)
   Digunakan untuk mengidentifikasi fitur paling berpengaruh terhadap harga (mileage, usia, brand, fuel_type).

B. Arsitektur Sistem Prediksi Harga Mobil Bekas

Arsitektur sistem menggambarkan bagaimana data input melalui proses preprocessing, kemudian diprediksi oleh model Random Forest Regressor, dan menghasilkan output berupa estimasi harga.



Gambar 3. Arsitektur Sistem Prediksi Harga Mobil Bekas

Penjelasan Arsitektur Sistem

1. Input Data (Fitur X)
   Sistem menerima input berupa atribut kendaraan, meliputi:
   a. milage_clean
   b. Usia_Mobil
   c. brand (hasil one-hot encoding)
   d. fuel_type
   e. transmission
   f. accident
2. Preprocessing Engine
   Semua input melewati modul pra-pemrosesan yang berfungsi untuk:
   a. membersihkan format data
   b. mengubah data kategorikal menjadi one-hot vectors
   c. memastikan data sesuai dengan struktur fitur pelatihan model
3. Model Random Forest Regressor (RFR)
   Model RFR memproses fitur input dan menggabungkan prediksi dari ratusan decision tree untuk menghasilkan nilai akhir.
   Sistem menggunakan model RFR hasil pelatihan terbaik yang memiliki:
   a. $R^2 = 0.68$
   b. MAE ≈ $12.2K
   c. RMSE ≈ $19.8K
4. Output Sistem
   Model menghasilkan output berupa:
   Prediksi harga mobil bekas (dalam USD) Output ditampilkan secara numerik dan dapat diintegrasikan dengan dashboard, aplikasi web, ataupun sistem internal dealer.

C. Rancangan Input dan Output Model

1. Rancangan Input (Fitur X)
   Input yang digunakan pada model prediksi merupakan kumpulan fitur numerik dan kategorikal yang telah melalui proses pembersihan, imputasi, dan One-Hot Encoding. Secara umum, fitur dibagi menjadi dua kelompok utama:

Tabel 1 Tabel Kategori Fitur

Jenis Fitur	Contoh Variabel
Numerik	*milage_clean*, *Usia_Mobil*
Kategorikal → One-Hot Encoding	*brand_*, *fuel_type_*, *transmission_*, *accident_*

Setelah seluruh variabel kategorikal dienkode menjadi fitur biner, total keseluruhan fitur input (X) yang digunakan pada model adalah 66 kolom. Struktur fitur ini memungkinkan model untuk menangkap pola non-linier serta interaksi antar variabel dalam proses prediksi harga mobil bekas.

2. Rancangan Output
   Output sistem adalah nilai prediksi harga mobil bekas, yaitu:

$$\hat{y} = \text{Predicted Used Car Price (USD)}$$

Output model dapat digunakan untuk:

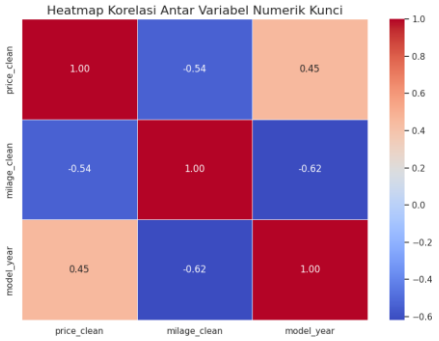
a. Estimasi harga jual–beli, baik untuk penjual maupun pembeli guna memperoleh perkiraan harga yang objektif.
b. Rekomendasi harga bagi dealer, sehingga proses penilaian kendaraan lebih konsisten dan tidak bergantung pada subjektivitas penaksir.
c. Mendukung keputusan konsumen, khususnya dalam menilai kewajaran harga, membandingkan alternatif kendaraan, dan menentukan nilai tawar yang lebih akurat.

Pemodelan

Bab ini menyajikan hasil eksperimen yang diperoleh dari proses eksplorasi data, pra-pemrosesan, pemodelan, hingga evaluasi model. Pembahasan difokuskan pada interpretasi hasil visualisasi EDA, performa kedua model (MLR dan Random Forest Regressor), serta analisis feature importance sebagai dasar penentuan faktor paling berpengaruh terhadap harga mobil bekas.

A. Hasil Eksplorasi Data (EDA)

1. Heatmap Korelasi Antar Variabel
   Gambar 4.6 menunjukkan korelasi antara variabel numerik utama, yaitu price_clean, milage_clean, dan model_year. Visualisasi ini dibuat setelah data outlier dihapus sehingga pola hubungan variabel lebih stabil dan representatif.
   Interpretasi Korelasi:
   a. Korelasi negatif antara mileage dan harga (r = −0.54)

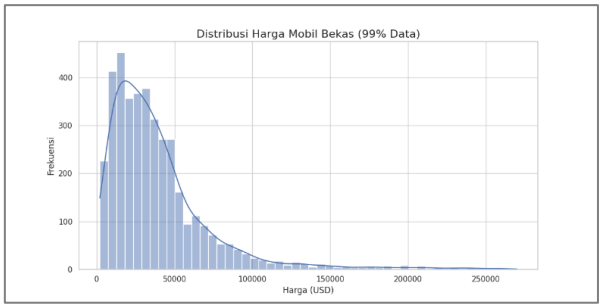


Gambar 3. Heatmap korelasi fitur numerik

Mengindikasikan bahwa semakin tinggi jarak tempuh mobil, semakin rendah harga jualnya. Hal ini sesuai dengan teori depresiasi aset kendaraan yang menurun seiring pemakaian [16], [17].

b. Korelasi positif antara tahun produksi dan harga (r = +0.45)
Mobil yang lebih baru memiliki harga lebih tinggi. Ini menggambarkan hubungan alami antara umur kendaraan dan nilai jual kembali.
c. Korelasi negatif antara mileage dan tahun produksi (r = −0.62)
Mobil yang lebih tua memiliki mileage yang lebih tinggi — pola umum di pasar otomotif.
d. Distribusi Harga Setelah Outlier Dihapus

Penghapusan 37 hyper-car (harga > $275.000) menghasilkan distribusi harga yang lebih normal dan tidak ekstrem.



Gambar 3. Distribusi Harga Setelah Outlier Dihapus

Kesimpulan EDA:

EDA menunjukkan bahwa harga mobil bekas dipengaruhi secara kuat oleh usia, jarak tempuh, dan faktor historis (accident). Pola-pola ini menjadi dasar pemilihan model non-linear seperti Random Forest.

B. Perbandingan Performa Model

Performa kedua model dievaluasi menggunakan R^2, MAE, MSE, dan RMSE untuk mengetahui akurasi serta stabilitas hasil prediksi.

Tabel 2. Perbandingan Kinerja Model MLR dan Random Forest

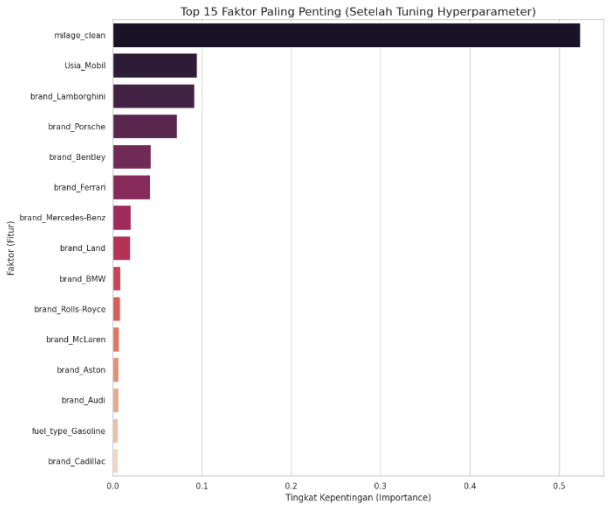
Model	R^2	MAE	MSE	RMSE
Regresi Linear	0.5673	$14,792.71	526,067,491.17	$22,936.16
Random Forest (untuned)	0.6827	$12,276.44	385,792,256.09	$19,641.60
Random Forest (tuned)	0.6751	$12,218.20	395,065,741.05	$19,876.26

Analisis Perbandingan:

1. Random Forest R^2 = 0.6827
Jauh lebih tinggi daripada MLR (0.5673), sehingga RFR mampu menjelaskan keragaman harga lebih baik.
2. Error lebih rendah pada RFR
MAE dan RMSE lebih kecil dibanding Regresi Linear → prediksi lebih dekat dengan nilai aktual.
3. Mengapa Random Forest lebih unggul?
    a. Mampu menangkap hubungan non-linear antara fitur dan harga.
    b. Dapat memodelkan interaksi antar variabel (misalnya brand × usia).
    c. Robust terhadap outlier ringan.
    d. Menghasilkan prediksi yang lebih stabil pada data heterogen (Gao, 2024; Aruna et al., 2021).
4. Tuning tidak selalu meningkatkan R^2
Penurunan kecil pada R^2 model tuned menunjukkan potensi *mild overfitting*, meski MAE membaik kondisi ini masih wajar pada model ensemble.

C. Analisis Feature Importance

Analisis feature importance dari Random Forest memberikan informasi mengenai fitur apa yang paling berpengaruh dalam penentuan harga mobil bekas.



Gambar 4. Lima Belas Fitur Paling Penting

Interpretasi Lima Fitur Terpenting:

1. milage_clean (paling penting)
   Mileage menjadi indikator kondisi kendaraan. Semakin tinggi angka ini, semakin besar depresiasi nilai mobil. Ini konsisten dengan teori depresiasi aset dan penelitian sebelumnya [3], [16].
2. Usia_Mobil (hasil rekayasa fitur)
   Usia mobil berbanding terbalik dengan nilai jualnya. Mobil berusia di atas 7 tahun cenderung mengalami penurunan harga lebih drastis.
3. Brand (brand_Toyota, brand_BMW, brand_Honda, dll.)
   Merek memengaruhi persepsi kualitas dan daya tahan kendaraan, sehingga berperan penting dalam valuasi harga.
4. fuel_type
   Mobil Hybrid/Electric cenderung memiliki harga lebih tinggi, selaras dengan tren kendaraan ramah lingkungan.
5. accident_None reported
   Mobil tanpa riwayat kecelakaan memiliki nilai yang lebih tinggi karena kondisi struktural lebih terjaga.

Kesimpulan dari Feature Importance:

a. Faktor teknis (mileage & usia) adalah indikator dominan harga mobil.
b. Faktor reputasi merek dan kondisi historis turut memperkuat prediksi.
c. Temuan ini memvalidasi teori depresiasi kendaraan serta hasil penelitian terdahulu [4], [11].

D. Perbandingan dengan Penelitian Terdahulu

Hasil penelitian menunjukkan bahwa Random Forest Regressor memiliki performa prediktif yang lebih baik dibandingkan Multiple Linear Regression (MLR), dengan nilai R^2 lebih tinggi dan error yang lebih rendah. Temuan ini selaras dengan berbagai studi terdahulu yang menyoroti keunggulan model ensemble untuk prediksi harga kendaraan dan aset lain. Ringkasan perbandingan ditampilkan pada Tabel 3.

Tabel 2 Perbandingan dengan Penelitian Terdahulu

Peneliti	**Metode**	**Dataset**	**Temuan Utama**	**Kesesuaian dengan Penelitian Ini**
Aruna et al. (2021)	Random Forest + GridSearchCV	Kaggle resale car	Tuning meningkatkan akurasi	Konsisten — tuning meningkatkan performa, meski rentan *overfitting* ringan
Gao (2024)	MLR vs RFR	Used cars (China)	RFR unggul untuk hubungan non-linear	Sejalan — RFR menangkap interaksi fitur lebih baik
Gollapalli et al. (2023)	RF, SVR, Decision Tree	Saudi used cars	RF paling akurat pada data heterogen	Selaras — RF stabil pada kombinasi data numerik & kategorikal

Umapathi & Latha (2024)	Random Forest	Pre-owned vehicle data	Mileage & usia = fitur paling dominan	Sama — mileage & usia menjadi fitur terpenting
Fitri (2023)	Linear, RF, GBT	Housing prices	Model ensemble paling akurat	Mendukung — ensemble stabil dan error lebih kecil
Penelitian ini (2025)	MLR vs RFR	Used Cars (Kaggle)	$R^2 = 0.68$; RMSE = 19,876	Memvalidasi keunggulan ensemble untuk valuasi kendaraan

Analisis Kritis dan Integratif

1. Keunggulan Ensemble Learning
   Mayoritas literatur [3], [8], [9], [11] menunjukkan bahwa Random Forest lebih efektif dibanding MLR untuk prediksi harga karena:
   a. mampu memodelkan non-linearitas,
   b. stabil terhadap outlier ringan,
   c. bekerja baik pada dataset heterogen,
   d. dapat menangkap interaksi antar fitur.
   Hasil penelitian ini yang memperoleh $R^2 \approx 0.68$, menunjukkan pola yang sama.
2. Dominasi Fitur Mileage dan Usia Kendaraan
   Penelitian [4], [17], dan [18] menegaskan bahwa usia mobil dan mileage adalah determinan utama harga. Feature importance dari penelitian ini juga menempatkan kedua variabel tersebut sebagai faktor paling berpengaruh konsistensi lintas studi.
3. Dampak Hyperparameter Tuning
   Sejalan dengan [8], tuning RF:
   a. meningkatkan MAE,
   b. namun sedikit menurunkan R^2,
   c. mengindikasikan *mild overfitting*.
   Fenomena ini umum pada ensemble apabila ruang parameter yang dicari terlalu sempit/luas.
4. Feature Importance sebagai Pendukung Interpretabilitas
   Studi [19] dan [20] menekankan bahwa *feature importance* meningkatkan transparansi model. Penelitian ini memperkuat temuan tersebut dengan menunjukkan:
   a. mileage_clean dan Car_Age sebagai fitur dominan,
   b. serta penggunaan RF yang memungkinkan interpretasi berbasis kontribusi variabel.

Kesimpulan

Penelitian ini bertujuan untuk mengembangkan model prediksi harga mobil bekas berbasis machine learning dengan membandingkan performa Regresi Linear dan Random Forest Regressor. Berdasarkan hasil eksperimen, analisis data, serta evaluasi model, diperoleh beberapa kesimpulan utama sebagai berikut:

1. Random Forest Regressor merupakan model terbaik dalam penelitian ini, dengan nilai akurasi tertinggi yaitu $R^2 = 0.6827$, melampaui model Regresi Linear yang hanya mencapai $R^2 = 0.5673$. Hal ini menunjukkan bahwa pendekatan ensemble learning lebih mampu menangkap pola non-linear dan interaksi kompleks antar fitur pada data mobil bekas.

2. Penerapan outlier filtering terhadap 37 data hyper-car (harga > $275.000) terbukti merupakan langkah krusial. Penghapusan outlier meningkatkan performa model secara signifikan—R^2 naik dari 0.11 menjadi 0.68. Hasil ini menegaskan bahwa kualitas data memiliki peran penting dalam keandalan model prediksi.

3. Faktor yang paling mempengaruhi harga mobil bekas adalah jarak tempuh (milage_clean) dan usia kendaraan (Usia_Mobil). Analisis feature importance menunjukkan bahwa kedua variabel tersebut merupakan indikator utama depresiasi kendaraan. Faktor lain seperti brand, fuel_type, dan accident history juga berkontribusi, namun tidak sekuat mileage dan usia.

4. Model yang dikembangkan mampu memberikan prediksi harga yang stabil dan dapat digunakan sebagai dasar evaluasi awal dalam sistem valuasi mobil bekas berbasis data. Temuan ini mendukung penggunaan pendekatan machine learning untuk meningkatkan objektivitas dan konsistensi dalam penentuan harga kendaraan di industri otomotif.

Daftar Pustaka

[1] Y. Chen, "Research on Machine Learning-based Prediction of Coffee Futures Prices," 2024.

[2] C. Li, "Machine Learning-Based Models for Accurate Car Prices Prediction," 2024.

[3] J. Gao, "Second-hand car price prediction based on multiple linear regression and random forest," *Theoretical and Natural Science*, vol. 52, no. 1, pp. 31–40, 2024, doi: 10.54254/2753-8818/52/2024ch0105.

[4] J. Umapathi and R. Latha, "Automated pricing predictions for pre-owned vehicles using random forest," *IJERST*, vol. 2, no. 2, pp. 21–26, 2024, doi: 10.63458/ijerst.v2i2.80.

[5] L. Bukvić, J. Pašagić Škrinjar, T. Fratrović, and B. Abramović, "Price prediction and classification of used-vehicles using supervised machine learning," *Sustainability*, vol. 14, no. 24, p. 17034, 2022, doi: 10.3390/su142417034.

[6] A. Alhakamy, A. Alhowaity, A. A. Alatawi, and H. Alsaadi, "Are used cars more sustainable? Price prediction based on linear regression," *Sustainability*, vol. 15, no. 2, p. 911, 2023, doi: 10.3390/su15020911.

[7] B. Cui, Z. Ye, H. Zhao, Z. Renqing, L. Meng, and Y. Yang, "Used car price prediction based on the iterative framework of XGBoost+LightGBM," *Electronics (Basel).*, vol. 11, no. 18, p. 2932, 2022, doi: 10.3390/electronics11182932.

[8] M. Aruna, M. Anjana, H. Chauhan, and R. Deepa, "Optimized hyperparameter tuned random forest regressor algorithm in predicting resale car value based on grid search method," *International Journal of Advanced Research in Science Communication and Technology*, pp. 106–113, 2021, doi: 10.48175/ijarsct-1217.

[9] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, vol. 4, no. 1, pp. 2723–1453, 2023, doi: 10.52158/jacost.491.

[10] L. Breiman, "Random Forests," 2001.

[11] M. Gollapalli, T. Alqahtani, D. Alhamed, M. Alnassar, A. Alajmi, and Y. Alali, "Intelligent modelling techniques for predicting used cars prices in Saudi Arabia," *Mathematical Modelling and Engineering Problems*, vol. 10, no. 1, pp. 139–148, 2023, doi: 10.18280/mmep.100115.

[12] S. Bergmann and S. Feuerriegel, "Machine learning for predicting used car resale prices using granular vehicle equipment information," *Expert Syst. Appl.*, vol. 263, p. 125640, Mar. 2025, doi: 10.1016/j.eswa.2024.125640.

[13] N. O. Idris and F. Pontoiyo, "Evaluasi Model Machine Learning untuk Prediksi Harga Mobil dengan Perbandingan Ensemble dan Regresi Linear," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 4, no. 1, pp. 129–143, Jan. 2025, doi: 10.70340/jirsi.v4i1.181.

[14] Md. A. Kadir, B. Nassar, S. M. Jahangir Alam, S. Ahmed, and Md. K. Syfullah, "Car Price Prediction: A Comparative Study of Machine Learning and Deep Learning Approaches," in *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/ECCE64574.2025.11013535.

[15] H. Lahuddin, Muh. R. A. Muliawan, K. Takemoto, H. Darwis, S. R. Jabir, and R. Adawiyah, "Cryptocurrency Prices Forecasting Using LSTM, CNN, Transformer, TCN, and Hybrid Model: A Deep Learning Approach," in *2025 9th International Conference On Electrical, Electronics And Information Engineering (ICEEIE)*, IEEE, Sep. 2025, pp. 1–6. doi: 10.1109/ICEEIE66203.2025.11252474.

[16] A. Eklund, A. Berndt, and S. Sandberg, "Orchestrating an experiential value proposition: The case of a Scandinavian automotive manufacturer," *European Business Review*, vol. 34, no. 5, pp. 624–641, 2022, doi: 10.1108/ebr-07-2021-0149.

[17] Y. Danylenko, "Analysis of the used passenger car market in Ukraine," *Economics Finances Law*, pp. 24–36, 2023, doi: 10.37634/efp.2023.9.6.

[18] V. K. Gohil, "Car Price Prediction by Machine Learning Approach," *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 09, no. 09, pp. 1–9, Sep. 2025, doi: 10.55041/IJSREM52586.

[19] R. Cheng, P. Haste, E. Levens, and J. Bergmann, "Feature importance for estimating rating of perceived exertion from cardiorespiratory signals using machine learning," *Front. Sports Act. Living*, vol. 6, 2024, doi: 10.3389/fspor.2024.1448243.

[20] Y. Gerstorfer, M. Hahn-Klimroth, and L. Krieg, "A Notion of Feature Importance by Decorrelation and Detection of Trends by Random Forest Regression," *Regression. Data Science Journal*, vol. 22, pp. 1–14, doi: 10.5334/dsj.