

Analisis Akurasi Algoritma Random Forest Regression dalam Memprediksi Harga Mobil Bekas

¹Zibran Brilliantama, ²Wahyu Adriyano, ³Adhitya Bagas Pratama

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

¹ zibranbrilian01@gmail.com, ² wahyuandriano0115@gmail.com, ³ bgsxprrtm@gmail.com

Abstract

The swift development of Machine Learning has opened up possibilities for enhancing decision-making in various sectors, including the assessment of pre-owned automobiles. Finding a just market value for a second-hand car can be difficult since it relies on many variables, including brand, model, year of manufacture, mileage, fuel type, transmission type, and engine details. This research intends to analyze how well the Random Forest Regression method performs in predicting the prices of used cars by examining the vehicle features found in the dataset. The methodology includes gathering data, preparing it, transforming categorical data, splitting the dataset into training and test groups, performing feature engineering, adjusting hyperparameters, building the model, and evaluating performance. The accuracy of the model is measured using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results of the experiments indicate that Random Forest Regression can deliver accurate price estimates while successfully recognizing intricate relationships between vehicle characteristics. As a result, the suggested model could facilitate objective, consistent, and data-driven pricing strategies in the market for used cars.

Keywords: *Random Forest Regression, Used Car Price Prediction, Machine Learning, Model Evaluation, Data Analytics.*

Abstrak

Kemajuan dalam teknologi Machine Learning telah memberikan dampak yang besar di banyak sektor, terutama dalam membantu proses penentuan harga kendaraan bekas. Proses menetapkan harga jual kendaraan bekas sering kali menghadapi kesulitan karena dipengaruhi oleh berbagai faktor kendaraan, seperti merek, jenis, tahun pembuatan, jarak tempuh, tipe bahan bakar, jenis transmisi, dan detail mesin. Penelitian ini bertujuan untuk menilai efektivitas algoritma Random Forest Regression dalam meramalkan harga kendaraan bekas berdasarkan atribut-atribut yang ada pada dataset. Langkah-langkah yang dilakukan dalam penelitian ini mencakup pengumpulan informasi, tahap preprocessing, perubahan atribut kategorikal, pemisahan data menjadi set pelatihan dan pengujian, penciptaan fitur tambahan, penyesuaian hyperparameter, pembangunan model, dan penilaian performa dengan menggunakan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan koefisien determinasi (R^2). Temuan dari penelitian ini menunjukkan bahwa model Random Forest Regression mampu memberikan estimasi harga kendaraan bekas dengan akurasi yang baik dan juga dapat mengungkap hubungan kompleks antara ciri-ciri kendaraan dan nilai jualnya. Oleh karena itu, model yang telah dikembangkan ini berpotensi digunakan sebagai alat bantu dalam pengambilan keputusan untuk menentukan harga kendaraan bekas dengan cara yang lebih objektif, konsisten, dan sesuai kondisi pasar.

Kata Kunci: *Random Forest Regression, Prediksi Harga Mobil Bekas, Machine Learning, Evaluasi Model, Analisis Data.*

A. PENDAHULUAN

Perkembangan teknologi informasi yang didorong oleh kemajuan dalam Machine Learning telah membawa dampak signifikan ke berbagai bidang, termasuk sektor otomotif. Salah satu aplikasinya adalah penggunaan algoritma pembelajaran mesin untuk memperkirakan harga jual mobil bekas dengan lebih tepat. Jika dibandingkan dengan metode penilaian tradisional yang banyak bergantung pada pengalaman penjual, pendekatan berbasis data dapat memberikan estimasi yang lebih adil dan konsisten.

Menentukan harga mobil bekas bukanlah tugas yang mudah karena nilai suatu kendaraan dipengaruhi oleh banyak faktor. Elemen-elemen seperti merek, model, tahun pembuatan, jarak tempuh, jenis bahan bakar, jenis transmisi, kapasitas mesin, serta kondisi kendaraan memiliki pengaruh yang berbeda terhadap harga jual. Banyaknya variabel yang perlu diperhitungkan menjadikan proses penetapan harga semakin rumit jika hanya mengandalkan observasi atau insting. Oleh karena itu, diperlukan suatu metode yang dapat digunakan untuk melakukan menganalisis keterkaitan antarvariabel tersebut secara terstruktur sehingga estimasi harga dapat dilakukan lebih akurat.

Dalam aktivitas jual beli kendaraan bekas, harga yang ditetapkan dapat berpengaruh besar pada keberhasilan transaksi. Harga yang ditetapkan terlalu tinggi dapat menurunkan minat konsumen yang menyebabkan kerugian bagi penjual. Situasi ini menyoroti pentingnya pendekatan berbasis data untuk mendukung pengambilan keputusan. Dengan memanfaatkan data penjualan sebelumnya, algoritma Machine Learning dapat mengidentifikasi pola hubungan antara karakteristik kendaraan dan harga pasar, sehingga mampu menghasilkan prediksi yang lebih valid.

Salah satu algoritma yang sering diterapkan dalam masalah prediksi ialah Random Forest Regression. Algoritma ini merupakan metode pembelajaran ensemble yang menciptakan sejumlah pohon keputusan lalu menggabungkan prediksi dari semua pohon tersebut untuk mendapatkan estimasi yang lebih konsisten dan tepat. Pendekatan ini tidak hanya efektif untuk menghadapi data dengan berbagai karakteristik kompleks, tetapi juga mampu mengurangi kemungkinan masalah overfitting yang sering muncul pada model pohon keputusan tunggal. Berbagai penelitian ini menunjukkan bahwa Random Forest memberikan hasil yang baik dalam berbagai situasi prediksi, termasuk untuk perkiraan harga kendaraan berdasarkan atributnya.

Studi ini menggunakan dataset mobil bekas yang berisi beragam informasi mengenai spesifikasi kendaraan sebagai dasar untuk model prediksi. Semua data diproses melalui beberapa tahap, termasuk preprocessing, transformasi data kategorikal, pembentukan fitur, normalisasi data, pembagian data menjadi set pelatihan dan pengujian, serta pengoptimalan hyperparameter dengan GridSearchCV. Langkah-langkah tersebut dilakukan agar model yang dihasilkan optimal dalam kemampuan prediksinya dan mampu menyesuaikan pola dan membuat prediksi pada data yang belum digunakan sebelumnya.

Untuk mengevaluasi kualitas model, penelitian ini menggunakan beberapa metrik, seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2). Penggunaan beragam metrik tersebut bertujuan memberikan gambaran yang lebih jelas menyeluruh mengenai akurasi model dalam memperkirakan harga mobil bekas.

Penelitian ini menggunakan kerangka kerja CRISP-DM (Cross Industry Standard Process for Data Mining) sebagai panduan pelaksanaan. Kerangka kerja ini dipilih karena menyediakan tahapan yang sistematis, dimulai dari pemahaman permasalahan, persiapan data, pembangunan model, sampai evaluasi hasil, sehingga sangat sesuai untuk penelitian yang berfokus pada pengembangan model prediksi berbasis data.

Melalui studi ini, diharapkan untuk memahami seberapa efektif algoritma Random Forest Regression dalam meramalkan harga mobil second berdasarkan fitur kendaraan yang terdapat dalam dataset. Selain itu, hasil

penelitian ini diharapkan dapat dijadikan sebagai acuan bagi pengembangan sistem serta pelaku industri otomotif dalam menciptakan sistem pendukung keputusan yang dapat memberikan estimasi harga kendaraan dengan lebih objektif, konsisten, dan sesuai dengan kondisi pasar.

B. METODE

1. Pengumpulan Data

Data yang digunakan adalah kumpulan harga mobil bekas yang terdiri dari 100 kendaraan dengan 13 fitur, yaitu Car_ID, Merek, Model, Tahun, Jarak_Tempuh, Jenis_Bahan_Bakar, Transmisi, Jenis_Pemilik, Konsumsi_BBM, Mesin, Daya, Jumlah_Kursi, dan Harga. Kumpulan data ini menyimpan rincian tentang karakteristik setiap kendaraan serta harganya. Semua fitur digunakan untuk menggambarkan keadaan kendaraan, sedangkan fitur Harga berfungsi sebagai variabel yang ingin diprediksi oleh model. Kumpulan data ini menjadi landasan dalam proses pembelajaran dan evaluasi model Regresi Hutan Acak.

2. Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan kualitas dan kesiapan data dalam kondisi yang siap sebelum digunakan dalam proses pembelajaran model. Proses ini dimulai dengan memeriksa struktur dataset, memeriksa jenis data, mengidentifikasi nilai yang hilang, serta memeriksa adanya data duplikat untuk menjaga kualitas dataset.

Selanjutnya, transformasi dilakukan terhadap atribut kategorikal seperti Brand, Model, Fuel_Type, Transmission, dan Owner_Type agar dapat dikelola oleh algoritma Machine Learning. Transformasi dilakukan menggunakan metode Label Encoding dan one hot Encoding sesuai dengan karakteristik masing-masing atribut. Atribut Car_ID tidak dipakai sebagai variabel prediktor, karena hanya berfungsi sebagai identitas unik untuk setiap data dan tidak berkontribusi dalam proses prediksi harga kendaraan.

Dari tahapan ini, diperoleh dataset yang lebih terstruktur dan bebas dari data yang tidak diperlukan., konsisten, serta telah siap digunakan dalam tahap pembuatan model.

Normalisasi Data

Setelah proses pra-pemrosesan selesai, dilakukan tahap normalisasi data untuk menyelaraskan rentang nilai antar atribut numerik sehingga proses pembelajaran model dapat berjalan lebih optimal. Pada penelitian ini digunakan metode StandardScaler yang mentransformasikan data ke dalam skala standar dengan rata-rata (mean) bernilai 0 dan simpangan baku (standard deviation) bernilai 1. Normalisasi diterapkan pada atribut numerik yang digunakan dalam pemodelan, seperti Kilometers_Driven, Mileage, Engine, Power, Car_Age, dan Power_per_CC. Tahap ini bertujuan untuk mengurangi perbedaan skala

antar fitur, meningkatkan konsistensi data, serta membantu model dalam mengidentifikasi pola hubungan antara karakteristik kendaraan dan harga mobil bekas secara lebih efektif. Meskipun algoritma Random Forest Regression tidak secara langsung bergantung pada skala data, proses standarisasi tetap dilakukan untuk mempertahankan kualitas data dan mendukung proses evaluasi model.

Penentuan Model dan Hyperparameter Optimal

Penentuan parameter optimal dilakukan menggunakan metode Grid Search dan Cross Validation untuk memperoleh kombinasi parameter terbaik pada model Random Forest Regression. Parameter yang diuji meliputi `n_estimators`, `max_depth`, `min_samples_split`, `min_samples_leaf`, dan `max_features`. Berdasarkan hasil evaluasi menggunakan nilai R^2 , MSE, dan RMSE, dipilih parameter dengan performa terbaik sehingga model dapat menghasilkan prediksi yang optimal

Pemodelan dan evaluasi dilakukan menggunakan algoritma Random Forest Regression dengan menerapkan metode Grid Search Cross Validation untuk memperoleh kombinasi parameter terbaik. Penilaian terhadap kinerja model dilakukan dengan menggunakan nilai R^2 Score, MSE, dan RMSE untuk mengukur kualitas hasil prediksi. Berdasarkan hasil pengujian tersebut, dipilih model dengan nilai R^2 tertinggi dan tingkat kesalahan prediksi terendah sehingga menghasilkan performa model yang optimal

C. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dari metode penelitian yang telah diuraikan pada bab sebelumnya untuk menganalisis kemampuan algoritma Random Forest Regression dalam memprediksi harga mobil bekas berdasarkan karakteristik kendaraan yang tersedia pada dataset. Seluruh tahapan penelitian dilaksanakan secara sistematis agar model yang dibangun mampu menghasilkan prediksi yang akurat dan memiliki kemampuan generalisasi yang baik. Proses tersebut diawali dengan preprocessing data untuk memastikan kualitas dan kelayakan data, dilanjutkan dengan normalisasi pada atribut numerik guna memperoleh representasi data yang lebih konsisten. Setelah itu dilakukan pembagian data menjadi data pelatihan dan data pengujian, kemudian dilanjutkan dengan proses penentuan kombinasi hyperparameter terbaik menggunakan metode GridSearchCV sebelum model Random Forest Regression dibangun. Tahap akhir penelitian berupa evaluasi performa model menggunakan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan koefisien determinasi (R^2 Score). Hasil dari setiap tahapan tersebut selanjutnya dibahas secara bertahap untuk memberikan gambaran mengenai proses pembentukan model serta tingkat kemampuan model dalam melakukan prediksi harga mobil bekas dengan memanfaatkan data yang tersedia sebagai dasar analisis.

Diproses secara langsung oleh algoritma Random Forest Regression. Hasil tersebut menunjukkan bahwa struktur dataset telah tersusun dengan baik dan dapat dilanjutkan ke tahapan pemeriksaan kualitas data.

Langkah berikutnya difokuskan pada identifikasi kualitas data melalui pengecekan missing value dan data duplikat. Hasil pemeriksaan menunjukkan bahwa seluruh atribut memiliki 100 nilai non-null, sehingga tidak ditemukan data yang hilang pada setiap kolom. Pemeriksaan kembali menggunakan fungsi `isnull()` juga menghasilkan nilai nol pada seluruh atribut, yang menandakan bahwa dataset telah lengkap dan tidak memerlukan proses imputasi data. Selain itu, hasil pengecekan data duplikat menunjukkan jumlah duplikat sebanyak 0, sehingga setiap baris data merepresentasikan observasi yang berbeda. Kondisi tersebut menunjukkan bahwa data yang digunakan memiliki tingkat konsistensi yang baik dan dapat langsung digunakan pada proses transformasi data tanpa memerlukan penanganan terhadap data kosong maupun data yang berulang.

Setelah kualitas data dipastikan memenuhi kebutuhan analisis, dilakukan proses transformasi terhadap atribut kategorikal agar seluruh fitur dapat diproses oleh model pembelajaran mesin. Atribut Brand dan Model dikonversi menjadi format numerik dengan menerapkan metode Label Encoding, sehingga setiap kategori direpresentasikan dalam bentuk kode bilangan bulat tanpa mengubah jumlah data yang tersedia. Sementara itu, atribut Fuel_Type, Transmission, dan Owner_Type ditransformasikan menggunakan teknik one-hot encoding dengan menghilangkan satu kategori dasar (`drop_first=True`) untuk menghindari terjadinya multikolinearitas pada data hasil transformasi. Setelah proses encoding selesai, pemeriksaan ulang terhadap struktur dataset memperlihatkan bahwa atribut yang sebelumnya bertipe object telah berubah menjadi tipe numerik sehingga seluruh fitur telah sesuai untuk digunakan pada tahap pemodelan. Perubahan ini juga menyebabkan penggunaan memori menjadi lebih efisien dibandingkan sebelum proses transformasi dilakukan.

Selain proses transformasi atribut kategorikal, pada tahap preprocessing juga dilakukan penyesuaian terhadap atribut yang digunakan dalam pemodelan. Atribut Car_ID tetap dipertahankan pada dataset hasil transformasi, namun tidak digunakan sebagai variabel prediktor karena hanya berfungsi sebagai identitas unik setiap data dan tidak memiliki hubungan langsung dengan harga mobil bekas yang diprediksi. Tahap preprocessing secara keseluruhan berhasil menghasilkan dataset yang lebih bersih, konsisten, dan seluruh fiturnya telah berada dalam format yang sesuai untuk memasuki tahapan berikutnya, yaitu normalisasi data dan pembangunan model Random Forest Regression. Dengan demikian, kualitas data yang telah dipersiapkan pada tahap ini diharapkan mampu mendukung proses pembelajaran model sehingga menghasilkan prediksi yang lebih akurat dan stabil.

1. Hasil Eksplorasi dan Preprocessing Data

Pemeriksaan awal menunjukkan bahwa dataset terdiri atas 100 data kendaraan dengan 13 atribut, yaitu Car_ID, Brand, Model, Year, Kilometers_Driven, Fuel_Type, Transmission, Owner_Type, Mileage, Engine, Power, Seats, dan Price. Struktur data terbaca dengan baik dan seluruh atribut berhasil dikenali oleh sistem. Pemeriksaan kualitas data memperlihatkan bahwa tidak terdapat missing value maupun data duplikat sehingga seluruh observasi dapat digunakan dalam proses analisis. Atribut kategorikal kemudian ditransformasikan ke bentuk numerik melalui Label Encoding dan One-Hot Encoding agar dapat diproses oleh algoritma Random Forest Regression. Selain itu, atribut Car_ID tidak digunakan sebagai variabel prediktor karena hanya berfungsi sebagai identitas data.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Car_ID          100 non-null   int64
1   Brand           100 non-null   object
2   Model           100 non-null   object
3   Year             100 non-null   int64
4   Kilometers_Driven 100 non-null   int64
5   Fuel_Type       100 non-null   object
6   Transmission    100 non-null   object
7   Owner_Type      100 non-null   object
8   Mileage         100 non-null   int64
9   Engine          100 non-null   int64
10  Power           100 non-null   int64
11  Seats           100 non-null   int64
12  Price           100 non-null   int64
dtypes: int64(8), object(5)
memory usage: 10.3+ KB
```

Gambar 1 Struktur Dataset

Berdasarkan Gambar 1, dataset terdiri atas 100 data dengan 13 atribut yang merepresentasikan karakteristik kendaraan serta harga jual mobil bekas. Pemeriksaan juga menunjukkan bahwa dataset memiliki atribut bertipe numerik (int64) dan atribut bertipe kategorikal (object). Keberadaan atribut kategorikal menunjukkan bahwa proses transformasi data diperlukan sebelum memasuki tahap pemodelan. Secara keseluruhan, struktur dataset telah berhasil dikenali oleh sistem sehingga dapat dilanjutkan ke tahap pemeriksaan kualitas data.

```
#Cek missing values
df.isnull().sum()

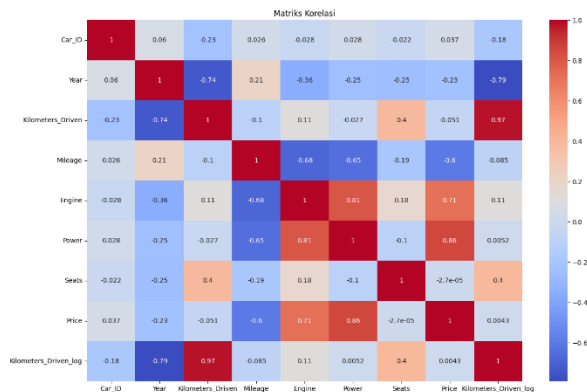
Car_ID      0
Brand       0
Model       0
Year        0
Kilometers_Driven 0
Fuel_Type   0
Transmission 0
Owner_Type  0
Mileage     0
Engine      0
Power       0
Seats       0
Price       0
```

Gambar 2 Hasil Pemeriksaan Data

Hasil pemeriksaan pada Gambar 2 menunjukkan bahwa dataset tidak memiliki missing value maupun data duplikat. Kondisi tersebut menunjukkan bahwa data telah memenuhi aspek kelengkapan dan konsistensi sehingga dapat digunakan pada tahap selanjutnya tanpa memerlukan proses pembersihan tambahan.

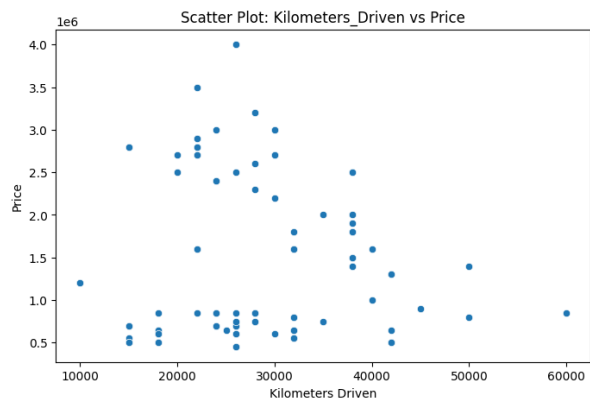
Gambar 3 Hasil Transformasi Data

Setelah kualitas data dipastikan baik, atribut kategorikal ditransformasikan menjadi format numerik dengan menerapkan teknik Label Encoding dan One-Hot Encoding. Selain itu, atribut Car_ID tidak digunakan sebagai variabel prediktor karena hanya berfungsi sebagai identitas data. Hasil transformasi tersebut menghasilkan dataset yang seluruh fiturnya telah sesuai untuk digunakan pada proses pembentukan model Random Forest Regression.



Gambar 4 Heatmap Korelasi Antar Atribut

Heatmap korelasi digunakan untuk memberikan gambaran awal mengenai hubungan antar atribut numerik pada dataset. Visualisasi ini membantu mengidentifikasi atribut yang memiliki hubungan terhadap variabel **Price**, sehingga dapat memberikan pemahaman awal mengenai karakteristik data sebelum dilakukan proses pemodelan.



Gambar 5 Scatter Plot Kilometers Driven vs Price

Berdasarkan Gambar 5, terlihat adanya kecenderungan bahwa semakin tinggi nilai **Kilometers_Driven**, harga mobil bekas cenderung mengalami penurunan. Meskipun demikian, penyebaran data masih menunjukkan variasi yang cukup besar pada setiap rentang jarak tempuh. Hal tersebut mengindikasikan bahwa harga mobil bekas tidak hanya dipengaruhi oleh jarak tempuh, tetapi juga oleh faktor lain seperti tahun kendaraan, kapasitas mesin, tenaga mesin, merek, dan spesifikasi lainnya. Visualisasi ini memberikan gambaran awal mengenai pola hubungan antara salah satu atribut dengan variabel target sebelum dilakukan proses pemodelan.

2. Pemodelan Random Forest Regression

Tahap pemodelan diawali dengan pembentukan fitur baru, yaitu **Car_Age** dan **Power_per_CC**, untuk memperkaya informasi yang digunakan dalam proses prediksi. Fitur **Car_Age** merepresentasikan usia kendaraan berdasarkan tahun produksi, sedangkan **Power_per_CC** menggambarkan rasio antara tenaga mesin dan kapasitas mesin. Penambahan kedua fitur tersebut bertujuan untuk membantu model dalam mengenali pola hubungan antara karakteristik kendaraan dan harga mobil bekas.

```
Best params : {'max_depth': 10, 'max_features': 0.5, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}
```

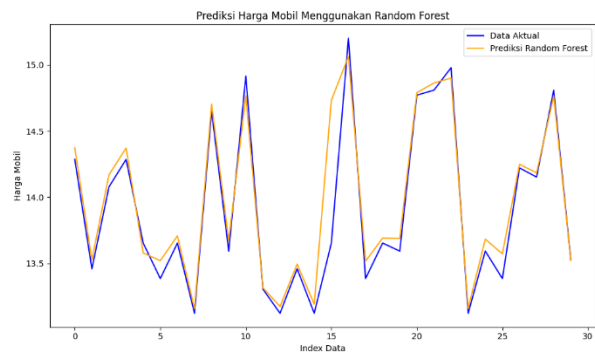
Gambar 6 Hasil Optimasi Hyperparameter Menggunakan GridSearchCV

Berdasarkan Gambar 6, proses optimasi hyperparameter menggunakan metode **GridSearchCV** berhasil mendapatkan kombinasi parameter yang paling optimal untuk digunakan dalam proses pembentukan model **Random Forest Regression**. Hasil optimasi menunjukkan bahwa model menggunakan nilai **max_depth = 10**, **max_features = 0.5**, **min_samples_leaf = 1**, **min_samples_split = 2**, dan **n_estimators = 200**. Konfigurasi tersebut menunjukkan bahwa model dibangun menggunakan 200 pohon keputusan (decision trees) dengan kedalaman maksimum 10 tingkat, sehingga mampu mempelajari pola hubungan antara karakteristik kendaraan dan harga mobil bekas tanpa menghasilkan model yang terlalu kompleks. Selain itu, nilai **max_features = 0.5** menunjukkan bahwa setiap pohon menggunakan sekitar 50% dari total fitur yang tersedia dalam proses pemilihan atribut terbaik pada setiap percabangan. Sementara itu, nilai **min_samples_split = 2** dan **min_samples_leaf = 1** memungkinkan model membentuk percabangan hingga batas minimum yang diperbolehkan, sehingga tetap mampu menangkap variasi pola pada data. Kombinasi hyperparameter tersebut kemudian digunakan sebagai konfigurasi akhir dalam proses pelatihan model untuk memperoleh performa prediksi yang optimal.

3. Evaluasi Model

Setelah model **Random Forest Regression** berhasil dibangun menggunakan kombinasi hyperparameter terbaik, tahap selanjutnya adalah mengevaluasi performa model terhadap data pengujian. Evaluasi dilakukan dengan membandingkan hasil prediksi dengan nilai aktual harga mobil bekas. Selain itu, performa model diukur

menggunakan beberapa metrik evaluasi, yaitu **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)**, dan **Coefficient of Determination (R² Score)**. Penggunaan beberapa metrik tersebut bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai tingkat akurasi model dalam melakukan prediksi.



Gambar 7 Perbandingan Data Aktual dan Hasil Prediksi Menggunakan Random Forest Regression

Berdasarkan Gambar 7, terlihat bahwa garis hasil prediksi (**Prediksi Random Forest**) mengikuti pola yang hampir sama dengan garis data aktual. Hal ini menunjukkan bahwa model mampu menangkap hubungan antara karakteristik kendaraan dan harga mobil bekas dengan baik. Meskipun pada beberapa titik masih terdapat sedikit perbedaan antara nilai prediksi dan nilai aktual, secara keseluruhan pola yang dihasilkan menunjukkan tingkat kesesuaian yang tinggi. Kedekatan kedua garis tersebut mengindikasikan bahwa model memiliki kemampuan yang baik dalam melakukan prediksi terhadap data pengujian.

```
Mean Squared Error (MSE): 0.0439  
Root Mean Squared Error (RMSE): 0.2095  
R2 Score: 0.8901  
R-squared (Train): 0.9951  
R-squared (Test): 0.8901
```

Gambar 8 Hasil Evaluasi Performa Model

Berdasarkan Gambar 8, model memperoleh nilai **Mean Squared Error (MSE)** sebesar 0,0439 dan **Root Mean Squared Error (RMSE)** sebesar 0,2095, yang menunjukkan bahwa tingkat kesalahan prediksi relatif rendah. Selain itu, model menghasilkan **R² Score** sebesar 0,8901, yang berarti sekitar 89,01% variasi harga mobil bekas dapat dijelaskan oleh model berdasarkan atribut yang digunakan dalam penelitian. Nilai **R² Train** sebesar 0,9951 dan **R² Test** sebesar 0,8901 menunjukkan bahwa model memiliki performa yang sangat baik pada data pelatihan dan tetap mampu mempertahankan akurasi yang tinggi ketika diuji menggunakan data yang belum pernah dipelajari sebelumnya. Secara keseluruhan, hasil evaluasi menunjukkan bahwa algoritma **Random Forest Regression** mampu memberikan prediksi harga mobil bekas dengan tingkat akurasi yang baik.

Implikasi Bisnis

Hasil penelitian menunjukkan bahwa algoritma Random Forest Regression mampu menghasilkan prediksi harga mobil bekas dengan tingkat akurasi yang baik berdasarkan karakteristik kendaraan yang digunakan dalam penelitian. Kemampuan model dalam memprediksi harga dapat dimanfaatkan sebagai sistem pendukung keputusan bagi pelaku usaha di bidang jual beli mobil bekas, seperti dealer, showroom, maupun platform marketplace. Dengan adanya estimasi harga yang lebih objektif, proses penentuan harga kendaraan dapat dilakukan secara lebih cepat dan konsisten sehingga mengurangi risiko penetapan harga yang terlalu tinggi (overpricing) maupun terlalu rendah (underpricing).

Selain membantu penjual dalam menentukan harga jual, model prediksi ini juga dapat memberikan manfaat bagi calon pembeli sebagai referensi dalam menilai kewajaran harga suatu kendaraan berdasarkan spesifikasinya. Penerapan model berbasis Random Forest Regression diharapkan mampu meningkatkan transparansi transaksi, mendukung pengambilan keputusan yang lebih berbasis data, serta meningkatkan efisiensi operasional dalam proses penilaian harga mobil bekas. Dengan demikian, hasil penelitian ini memiliki potensi untuk diimplementasikan sebagai bagian dari sistem informasi penentuan harga kendaraan yang lebih akurat dan mendukung daya saing bisnis di sektor otomotif.

D. PENUTUP

Simpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Random Forest Regression dapat diterapkan untuk melakukan prediksi terhadap harga mobil bekas berdasarkan karakteristik kendaraan, seperti tahun produksi, jarak tempuh, kapasitas mesin, tenaga mesin, jenis bahan bakar, tipe transmisi, dan atribut lainnya. Tahapan penelitian yang meliputi preprocessing data, transformasi atribut kategorikal, normalisasi data, optimasi hyperparameter menggunakan GridSearchCV,

serta pembangunan model berhasil diterapkan dengan baik sehingga menghasilkan model prediksi yang optimal.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan lebih beragam agar tingkat akurasi prediksi dapat ditingkatkan. Selain itu, penelitian berikutnya juga dapat mengembangkan metode atau menambahkan algoritma lain untuk memperoleh ditambahkan atribut lain yang memengaruhi harga mobil bekas serta dilakukan perbandingan dengan algoritma Machine Learning lainnya untuk memperoleh performa model yang lebih baik. Pengembangan model ke dalam aplikasi berbasis web atau mobile juga dapat dilakukan agar dapat dimanfaatkan secara langsung oleh pengguna.

E. DAFTAR PUSTAKA

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth International Group.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.