

PENGEMBANGAN SISTEM PREDIKSI HARGA MOBIL BEKAS OLX MENGGUNAKAN ALGORITMA RANDOM FOREST

Ervier Priande^{a,1,*} Guntur Maulana Zamroni^{b,2*}

^ac Jl Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55191

¹ ervier1800018009@webmail.uad.ac.id; ² guntur.zamroni@tif.uad.ac.id

* Penulis Korespondensi

ABSTRAK

Mobil merupakan kendaraan beroda empat yang populer karena kenyamanannya dalam bepergian dan perlindungan dari cuaca ekstrem. Dalam pasar mobil bekas yang berkembang, prediksi harga mobil bekas secara manual kerap tidak akurat karena keterbatasan data, waktu, kesalahan manusia, dan ketidakmampuan beradaptasi dengan perubahan pasar. Penelitian ini bertujuan membantu pengusaha mobil bekas memprediksi harga dengan lebih akurat dan efisien menggunakan algoritma Random Forest. Penelitian ini terdiri dari beberapa tahapan. Tahapan pertama yaitu mengumpulkan data produk kendaraan dari olx.com menggunakan webscraping Python dilanjutkan dengan preprocessing, analisis deskriptif, label encoding, menentukan variabel independen dan dependen dan pembagian data set untuk pengujian. Selanjutnya dilakukan konfigurasi model untuk melakukan beberapa pengujian dengan paramater yang berbeda. Setelah model pengujian prediksi terbentuk dilakukan penyimpanan model ke dalam format pickle serta melakukan perbandingan harga prediksi dengan harga sebenarnya. Hasil analisis menunjukkan bahwa merek mobil Toyota mendominasi dengan jumlah 580 unit, diikuti oleh Honda (348) dan Suzuki (153). Sebagian besar mobil menggunakan bahan bakar bensin (1684) daripada solar (167). Transmisi otomatis lebih umum (1462) dibandingkan dengan transmisi manual (389). Hasil prediksi harga mobil bekas menggunakan Algoritma Random forest menghasilkan tingkat akurasi sebesar 94,79% dan Mean Absolute Percentage Error (MAPE) sekitar 5.21% dan Mean Absolute Error (MAE) 9729279.74 dengan penggunaan 250 pohon(n_estimator), dengan perbedaan rata-rata antara harga aktual dan prediksi sekitar 0,10%.

Kata Kunci
Prediksi harga,
random forest,
webscraping



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

1. Pendahuluan

Industri otomotif saat ini merupakan salah satu industri global yang paling signifikan permintaan mobil terus meningkat setiap hari. Untuk mendapatkan dari satu tempat ke tempat lain, orang mengendarai mobilnya sendiri. Yang paling berguna dan mudah beradaptasi adalah kendaraan roda empat yaitu mobil. Selain itu, Pertumbuhan ekonomi suatu negara sangat bergantung pada transportasinya infrastruktur. Karena infrastruktur transportasi yang efektif memberikan manfaat ekonomi dan sosial peluang yang menguntungkan pasar [1].

Pelanggan lebih memilih memiliki mobil bekas karena jaminan yang diberikan oleh showroom dimana tempat mereka membeli mobil bekas tersebut dibandingkan melakukan pembelian mobil baru Karena biayanya mengeluarkan sejumlah besar uang untuk membeli mobil baru dari sudut pandang ekonomi [1].

Mobil adalah kendaraan beroda empat yang digerakkan oleh mesin dengan bahan bakar bensin atau solar dan memiliki bentuk tertentu. Mobil menjadi salah satu alat transportasi yang sangat populer di kalangan masyarakat karena memungkinkan seseorang untuk bepergian ke berbagai tempat dengan nyaman serta terlindungi dari kondisi cuaca yang ekstrem seperti hujan atau panas

terik. Ada berbagai jenis mobil yang tersedia, termasuk mobil Sport, SUV, Convertible, Coupe, Hatchback, MPV, Station Wagon, Sedan, dan Off-Road.

Dalam seiring perkembangan positif pada penjualan mobil bekas, semakin banyak variasi merek, model, dan harga yang tersedia. Prediksi harga mobil bekas dengan cara lama dapat dilakukan dengan melihat beberapa faktor utama yang dapat mempengaruhi nilai jual mobil bekas:

1. Tahun produksi mobil, semakin tua mobil, biasanya semakin rendah nilai jualnya. Namun, jika mobil tersebut termasuk mobil langka atau koleksi, maka nilainya bisa jadi lebih tinggi dari mobil sejenis yang lebih baru,
2. Kilometer mobil, semakin banyak jarak yang sudah ditempuh oleh mobil, biasanya semakin rendah nilai jualnya. Namun, jika mobil tersebut dijaga dengan baik dan telah mendapatkan perawatan berkala secara teratur, nilai jualnya bisa jadi lebih tinggi dari mobil sejenis yang memiliki jarak tempuh yang lebih rendah.
3. Merek mobil: Beberapa merek mobil memiliki reputasi yang lebih baik daripada merek lainnya, yang dapat mempengaruhi nilai jual mobil tersebut. Misalnya, merek mobil terkenal seperti Toyota atau Honda biasanya memiliki nilai jual yang lebih tinggi daripada merek mobil yang kurang terkenal.
4. Kondisi mobil: Kondisi mobil secara keseluruhan juga dapat mempengaruhi nilai jualnya. Jika mobil terlihat bersih dan terawat dengan baik, maka nilai jualnya bisa jadi lebih tinggi. Sebaliknya, jika mobil terlihat kotor dan rusak, nilai jualnya akan cenderung lebih rendah.
5. Fitur mobil: Beberapa fitur tambahan pada mobil seperti sistem audio, sunroof, atau kursi kulit, dapat meningkatkan nilai jualnya. Namun, jika fitur-fitur tersebut sudah usang atau tidak berfungsi dengan baik, maka nilai jualnya bisa jadi lebih rendah [2].

Setelah mengetahui faktor-faktor yang mempengaruhi nilai jual mobil bekas, maka dapat mencari informasi tentang harga pasar mobil sejenis dengan melakukan riset online atau melalui teman atau kenalan yang berpengalaman dalam membeli atau menjual mobil bekas. Dari sana, maka akan dapat menghitung perkiraan harga mobil bekas yang ingin dijual atau dibeli dengan mempertimbangkan faktor-faktor tersebut. Namun, dengan cara tersebut bahwa prediksi harga mobil bekas secara manual tidak dapat memberikan hasil harga yang akurat.

Hal tersebut dapat diperkuat dari hasil pra-wawancara oleh bapak Arya selaku owner showroom nugraha motor, beliau menyatakan bahwa permasalahan yang mungkin muncul saat melakukan prediksi harga mobil bagi konsumen antara lain:

1. Keterbatasan data dan informasi: Dalam melakukan prediksi harga mobil, calon pembeli hanya dapat menggunakan data dan informasi yang terbatas untuk menentukan harga mobil. Hal ini dapat menyebabkan ketidakakuratan dalam memprediksi harga.
2. Keterbatasan waktu: Melakukan prediksi harga mobil secara manual membutuhkan waktu yang cukup lama dan memakan banyak waktu, terutama jika dilakukan secara terus-menerus. Keterbatasan waktu ini dapat menghambat efisiensi dalam menjual mobil dan mengakibatkan hilangnya kesempatan untuk menawarkan harga yang lebih baik kepada konsumen.
3. Kesalahan manusia: Prediksi harga mobil secara manual dapat dipengaruhi oleh faktor kesalahan manusia seperti kesalahan dalam menghitung biaya, kesalahan dalam menentukan faktor-faktor yang mempengaruhi harga mobil, atau kesalahan dalam menganalisis data yang tersedia. Hal ini dapat menyebabkan ketidakakuratan dalam memprediksi harga dan menyebabkan kehilangan pelanggan.
4. Tidak dapat menyesuaikan dengan perubahan pasar: Perkembangan pasar yang cepat membuat prediksi harga mobil secara manual tidak dapat menyesuaikan dengan cepat. Perubahan permintaan, persaingan, atau faktor-faktor lainnya dapat memengaruhi harga mobil secara signifikan.

Oleh karena itu, penulis melakukan penelitian untuk membantu seorang pengusaha jual beli mobil bekas dalam memprediksi harga mobil bekas berdasarkan kriteria seperti merek, model, jenis bahan bakar, jenis transmisi, kilo meter yang telah ditempuh, serta tahun pembuatan mobil menggunakan algoritma random forest agar prediksi tersebut lebih akurat dan efisiensi.

Dalam upaya memprediksi harga mobil bekas, diperlukan suatu algoritma yang tepat. Salah satu algoritma yang digunakan dalam penelitian ini adalah Random forest dipilih karena mampu mengatasi masalah data yang berjumlah besar dan efektif dalam menangani masalah data yang hilang

[3] . Algoritma Random forest juga termasuk dalam metode Classification and Regression Tree (CART), yang dapat digunakan untuk klasifikasi maupun prediksi, tanpa memerlukan asumsi tertentu. Akurasi dari algoritma Random forest dapat ditingkatkan dengan menggunakan metode ensemble, yaitu dengan menggabungkan banyak model CART.

Random Forest adalah metode yang dikembangkan dari Classification and Regression Tree (CART), yang mengadopsi teknik bagging atau bootstrap aggregating serta pemilihan fitur secara acak [4]. Metode ini menggunakan dasar pohon keputusan (decision tree), sehingga Random Forest terdiri dari banyak pohon keputusan yang membentuk sebuah hutan acak. Hutan acak ini berfungsi untuk mengklasifikasi atau memprediksi data tertentu. Dalam penelitian ini, variabel dependen yang dianalisis adalah harga, sementara variabel independennya mencakup merek, model, tahun, transmisi jenis bahan bakar dan kilo meter. Dengan memanfaatkan Random Forest, analisis dapat dilakukan dengan lebih akurat karena metode ini mampu mengatasi kompleksitas data dan variabilitas yang ada dalam variabel-variabel tersebut. Teknik ini memberikan keunggulan dalam menangani data yang beragam dan memungkinkan prediksi yang lebih andal dan tepat.

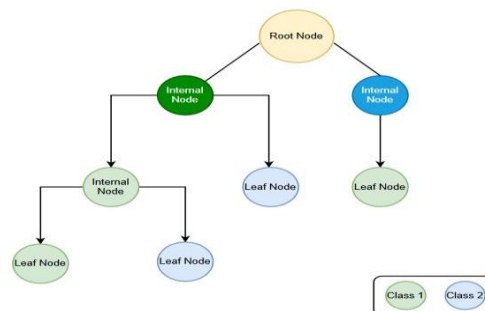
2. Metode

2.1 Web scraping

Web scraping merupakan teknik yang sangat penting dalam pengumpulan data dalam berbagai bidang, seperti riset pasar, analisis saham, dan penelitian akademis. Namun, teknik ini juga sering menjadi kontroversi karena dapat melanggar hak cipta, privasi, dan kebijakan penggunaan data dari sebuah website. Dalam implementasinya, web scraping memanfaatkan teknologi yang dikenal dengan crawler atau spider. Crawler merupakan program yang dapat mengunjungi halaman web secara otomatis dan mengambil data yang diinginkan. Kemudian data tersebut diolah untuk diambil informasi yang relevan dan disimpan dalam format yang sesuai [5].

2.2 Classification and Regression Tree (CART)

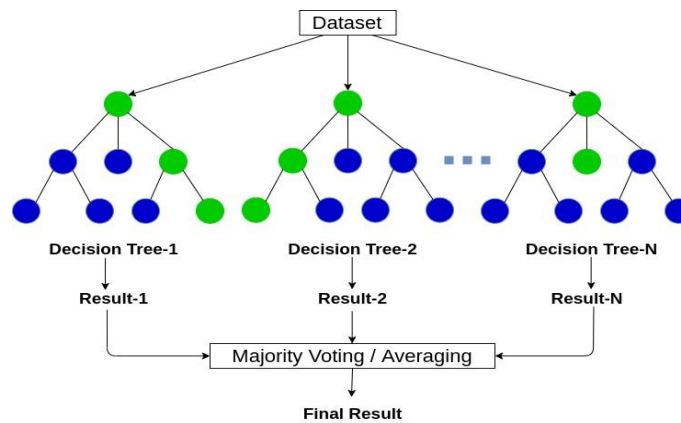
Classification and Regression Tree (CART) adalah metode dalam pembelajaran mesin yang digunakan untuk membuat model prediktif berdasarkan data terstruktur. Metode CART menghasilkan pohon keputusan yang terdiri dari serangkaian aturan keputusan untuk memprediksi nilai variabel target berdasarkan nilai variabel input. CART dapat digunakan untuk masalah klasifikasi (classification) dan regresi (regression) [6].



Gambar 1 Classification and Rgression Tree

2.3 Random forest

Random forest merupakan metode dalam pembelajaran mesin yang digunakan untuk membuat model prediktif dengan menggabungkan beberapa pohon keputusan (decision tree) yang saling tidak terkait. Metode ini dilakukan dengan mengambil sampel secara acak dari dataset dan membangun beberapa pohon keputusan pada setiap sampel. Setiap pohon keputusan pada Random forest akan memprediksi nilai target dan kemudian menghitung akurasi model prediktif dengan menggabungkan prediksi dari semua pohon keputusan [7].

Gambar 2 *Random forest*

2.4 Pengujian Akurasi

Menurut Carlo Vercellis dalam penelitian yang dikutip oleh [8], ada dua alasan utama mengapa pengukuran nilai akurasi model prediksi sangat penting. Pertama, nilai akurasi digunakan sebagai tolak ukur untuk membandingkan berbagai model prediksi alternatif. Dengan memiliki nilai akurasi, kita bisa menentukan model mana yang memberikan hasil paling akurat dibandingkan model lainnya. Kedua, akurasi juga berguna untuk mendeteksi kelemahan dalam model yang sudah dibuat. Jika akurasi rendah, kita bisa meneliti lebih lanjut untuk memperbaiki atau mengoptimalkan model tersebut.

Kinerja dari model prediksi yang dihasilkan melalui berbagai metode analisis bisa dievaluasi berdasarkan tingkat kesalahan atau error dari hasil prediksi. Salah satu cara untuk mengukur error adalah dengan menggunakan *Mean Absolute Percentage Error (MAPE)*. MAPE merupakan rata-rata persentase dari selisih antara nilai aktual dan nilai prediksi. Dengan kata lain, MAPE menunjukkan seberapa besar rata-rata kesalahan prediksi dalam bentuk persentase.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan melalui teknik scraping, yaitu pengambilan seluruh data dari sebuah situs web tertentu hingga batas waktu yang telah ditentukan. Data yang digunakan adalah data primer yang terdiri dari teks dan angka mengenai spesifikasi mobil yang terdapat pada situs web O LX.co.id. Untuk melakukan teknik pengambilan data, digunakan Bahasa pemrograman Python dan library Chrome Driver Manager. Pengambilan sampel data dilakukan dengan metode purposive sampling yang mempertimbangkan kriteria dan batasan tertentu yang sesuai dengan tujuan penelitian [9].

3.2. Implementasi

3.2.1 *Web scraping*

Dalam penelitian ini, data dikumpulkan menggunakan teknik *web scraping* menggunakan Python dan pustaka *ChromeDriverManager*. Hasil *web scraping* menghasilkan sebanyak 1856 data yang kemudian disimpan dalam format CSV pada file Excel.

id	model	nama	jenis	harga	tahun	bahan bakar	KM	transmisi
850748974	toyota	fortuner	,2.5 G Solar	Rp 188.000.000	2009	Diesel	105173 Km	Manual
853082501	honda	cr-v	,2.0 RE1 Bensin	Rp 139.000.000	2010	Bensin	101354 Km	Otomatis
824825116	bmw	520i	,2.0 Luxury Bensin	Rp 769.000.000	2019	Bensin	37417 Km	Otomatis
852897178	toyota	kijang	,1.8 SGX Bensin	Rp 69.000.000	2000	Bensin	250.000-255.000 Km	Manual
852086819	toyota	kijang innova	,2.0 Bensin	Rp 277.000.000	2017	Bensin	40.000-45.000 Km	Automatic
852270343	toyota	lexus	,2.7 RX270 Bensin	Rp 425.000.000	2014	Bensin	105.000-110.000 Km	Automatic
856159236	honda	mobilio	,1.5 E Bensin	Rp 120.000.000	2014	Bensin	25.000-30.000 Km	Automatic
852460888	honda	jazz	,1.5 S Bensin	Rp 197.000.000	2019	Bensin	0-5.000 Km	Manual
853743010	bmw	x1	,2.0 sDrive18i Bensin	Rp 215.000.000	2012	Bensin	95.000-100.000 Km	Automatic
853743261	toyota	lexus	,2.7 RX270 Bensin	Rp 330.000.000	2011	Bensin	35.000-40.000 Km	Automatic
853957561	mercedes-benz	gla200	,2.0 Bensin	Rp 470.000.000	2015	Bensin	15.000-20.000 Km	Automatic
853957682	honda	cr-z	,1.5 Hybrid	Rp 355.000.000	2013	Bensin	50.000-55.000 Km	Automatic
853957801	mitsubishi	pajero sport	,2.4 Dakar 4x2 Ultimate Solar	Rp 475.000.000	2018	Diesel	10.000-15.000 Km	Automatic
854036455	hyundai	h-1	,2.5 CRDi XG Solar	Rp 285.000.000	2015	Diesel	65.000-70.000 Km	Automatic
855889686	toyota	fortuner	,2.4 4x2 VRZ Solar	Rp 435.000.000	2018	Diesel	70.000-75.000 Km	Automatic
854644237	honda	mobilio	,1.5 RS Bensin	Rp 155.000.000	2016	Bensin	95.000-100.000 Km	Automatic
855889658	honda	city	,1.5 E Bensin	Rp 179.000.000	2015	Bensin	105.000-110.000 Km	Automatic
855889745	toyota	fortuner	,2.4 4x2 VRZ Solar	Rp 435.000.000	2018	Diesel	70.000-75.000 Km	Automatic
855889609	toyota	avanza	,1.3 Veloz Bensin	Rp 122.000.000	2016	Bensin	225.000-230.000 Km	Automatic
856170917	toyota	avanza	,1.3 E Bensin	Rp 120.000.000	2016	Bensin	160.000-165.000 Km	Manual
844933557	suzuki	ignis	,1.2 GL Bensin	Rp 92.000.000	2017	Bensin	83988 Km	Manual
855726562	toyota	venturer	,2.4 Solar	Rp 475.000.000	2019	Diesel	25.000-30.000 Km	Automatic
854148144	honda	iazz	,1.5 RS Bensin	Rp 219.000.000	2017	Bensin	8983 Km	Manual

Gambar 3 Hasil Web scraping

3.2.2 Preprocessing

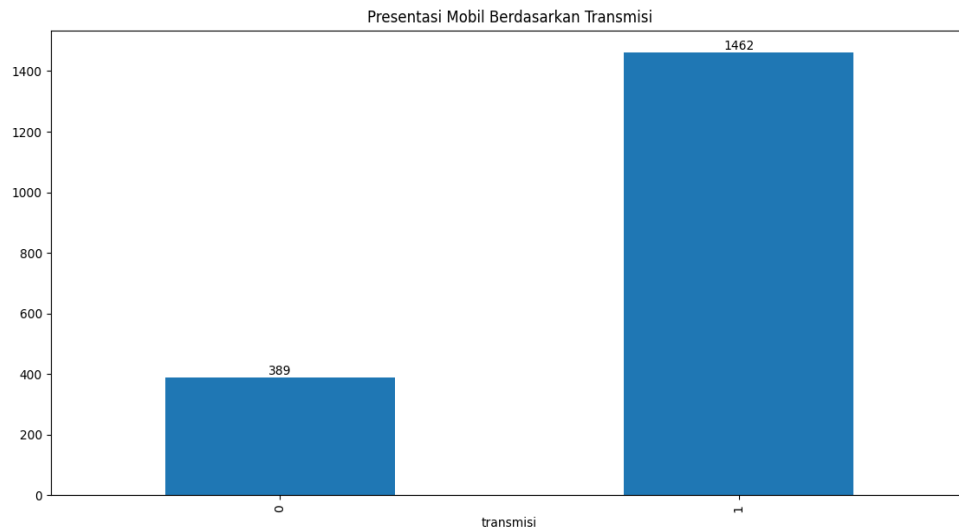
Setelah melakukan *web scraping*, maka perlu dilakukan data *preprocessing*. Dengan transformasi data untuk mengubah variabel kilometer (Km) menjadi beberapa kategori agar data lebih mudah dipahami. Informasi mengenai kategori kilometer dapat ditemukan pada Tabel 1.

Tabel 1 Kategori kilometer (Km)

Km	Kategori
0 – 10.000	0
10.000-20.000	1
20.000-30.000	2
30.000-40.000	3
40.000-50.000	4
50.000-60.000	5
60.000-70.000	6
70.000-80.000	7
80.000-90.000	8
90.000-100.000	9
> 100.000	10

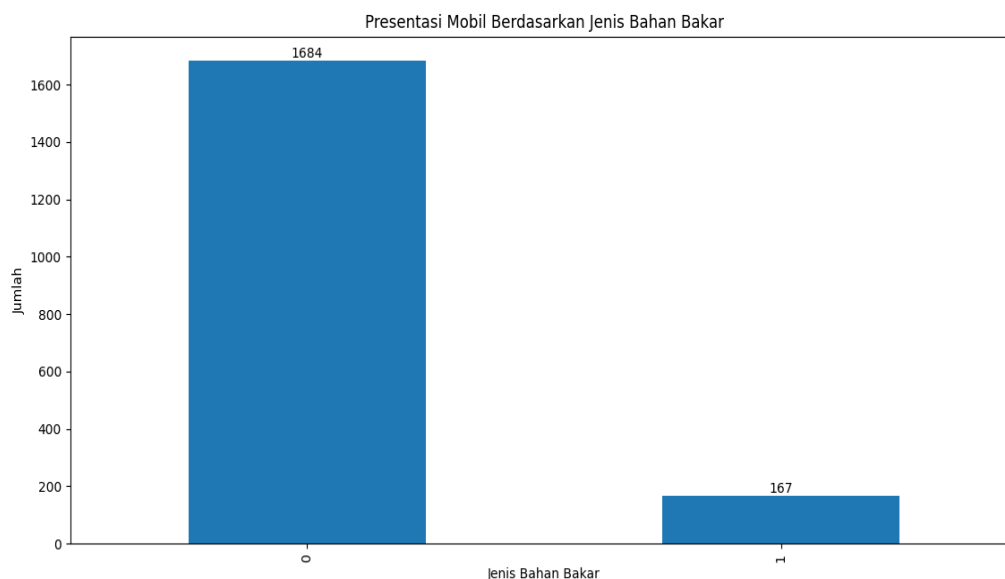
3.2.3 Analisa Deskriptif

Analisis deskriptif dalam penelitian ini dilakukan untuk memberikan gambaran umum tentang data yang dikumpulkan melalui *Web scraping* dari situs OLX.co.id. Berdasarkan hasil pengolahan data *Web scraping* menunjukkan bahwa transmisi yang paling banyak digunakan adalah otomatis dengan inisial 1 pada grafik dengan jumlah sebanyak 1462 mobil sedangkan transmisi manual dengan inisial 0 pada grafik hanya berjumlah 389 mobil, dapat dilihat pada Gambar 2



Gambar 4 Persentase Mobil Berdasarkan Jenis Transmisi

Berdasarkan grafik diatas pada gambar 4.3 menunjukkan bahwa sebanyak 1462 mobil menggunakan transmisi otomatis dan 389 mobil menggunakan transmisi manual. Mobil dengan transmisi otomatis lebih banyak digunakan pada masyarakat dibandingkan dengan transmisi manual itu disebabkan Transmisi otomatis lebih mudah digunakan, karena pengemudi tidak perlu terus-menerus mengganti gigi secara manual. Hal ini membuat pengemudi dapat fokus lebih pada mengemudi dan mengurangi kelelahan, terutama saat berkendara dalam lalu lintas padat atau dalam perjalanan jarak jauh serta Transmisi otomatis cenderung memberikan performa yang lebih konsisten karena sistemnya secara otomatis menyesuaikan perpindahan gigi dengan kondisi jalan dan kecepatan kendaraan. Hal ini dapat meningkatkan keamanan dan kenyamanan berkendara



Gambar 5 Persentase Mobil Berdasarkan Jenis Bahan bakar

Berdasarkan grafik diatas pada gambar 4.4 menunjukkan bahwa sebanyak 1684 mobil menggunakan bahan bakar bensin dan 167 mobil menggunakan bahan bakar solar. Mobil berbahan bakar bensin biasa menggunakan Pertalite, Pertamina, dan Pertamina Turbo sedangkan mobil berbahan bakar solar biasanya menggunakan bahan bakar Biosolar atau Pertamina Dex. Dari diagram tersebut dapat disimpulkan bahwa mobil dengan bahan bakar

bensin paling banyak dijual karena lebih mudah perawatannya serta lebih umum dimiliki masyarakat.

3.2.4 Label Encoding

Setelah itu dilakukan label *encoding* pada variabel lainnya yang masih berupa *string* menjadi *float* dengan cara sebagai berikut.

4	1,bmw,6,520i ,142,2.0 Luxury Bensin,769000000.0,2019,3.0,0,Bensin,3,1,Otomatis				
5	24,toyota,101,kijang ,127,1.8 SGX Bensin,69000000.0,2000,22.0,0,Bensin,10,0,Manual				
6	24,toyota,102,kijang innova ,131,2.0 Bensin,277000000.0,2017,5.0,0,Bensin,6,1,Otomatis				
7	24,toyota,107,lexus ,224,2.7 RX270 Bensin,425000000.0,2014,8.0,0,Bensin,10,1,Otomatis				
8	6,honda,112,mobilio ,62,1.5 E Bensin,120000000.0,2014,8.0,0,Bensin,10,1,Otomatis				
9	6,honda,96,jazz ,85,1.5 S Bensin,197000000.0,2019,3.0,0,Bensin,0,0,Manual				
10	1,bmw,157,x1 ,159,2.0 sDrive18i Bensin,215000000.0,2012,10.0,0,Bensin,10,1,Otomatis				
11	24,toyota,107,lexus ,224,2.7 RX270 Bensin,330000000.0,2011,11.0,0,Bensin,5,1,Otomatis				
12	15,mercedes-benz,81,gla200 ,131,2.0 Bensin,470000000.0,2015,7.0,0,Bensin,2,1,Otomatis				
13	6,honda,54,cr-z ,73,1.5 Hybrid,355000000.0,2013,9.0,0,Bensin,7,1,Otomatis				
14	17,mitsubishi,119,pajero sport ,172,2.4 Dakar 4x2 Ultimate Solar,475000000.0,2018,4.0,1,Diesel,1,1,Otomatis				
15	7,hyundai,89,h-1 ,197,2.5 CRDi XG Solar,285000000.0,2015,7.0,1,Diesel,10,1,Otomatis				
16	24,toyota,76,fortuner ,166,2.4 4x2 VRZ Solar,435000000.0,2018,4.0,1,Diesel,10,1,Otomatis				
17	6,honda,112,mobilio ,82,1.5 RS Bensin,155000000.0,2016,6.0,0,Bensin,10,1,Otomatis				
18	6,honda,41,city ,62,1.5 E Bensin,179000000.0,2015,7.0,0,Bensin,10,1,Otomatis				
19	24,toyota,76,fortuner ,166,2.4 4x2 VRZ Solar,435000000.0,2018,4.0,1,Diesel,10,1,Otomatis				
20	24,toyota,23,avanza ,44,1.3 Veloz Bensin,122000000.0,2016,6.0,0,Bensin,10,1,Otomatis				
21	24,toyota,23,avanza ,35,1.3 E Bensin,120000000.0,2016,6.0,0,Bensin,10,0,Manual				
22	23,suzuki,94,ignis ,17,1.2 GL Bensin,92000000.0,2017,5.0,0,Bensin,8,0,Manual				
23	24,toyota,151,venturer ,182,2.4 Solar,475000000.0,2019,3.0,1,Diesel,10,1,Otomatis				
24	6.honda.96.iazz .82.1.5 RS Bensin.219000000.0.2017.5.0.0.Bensin.0.0.Manual				

Gambar 6 Hasil Label *Encoding*

3.3. Pengujian

3.3.1 Menentukan Variabel Dependen dan Independen

Dalam tahap analisis algoritma *Random forest*, langkah pertama yang perlu dilakukan adalah menentukan variabel independen dan dependen. Dalam penelitian ini, terdapat 7 variabel independen yang meliputi merek, nama, model, tahun, jenis bahan bakar, kilometer (Km) jarak tempuh, dan jenis transmisi. Sementara itu, variabel dependen dalam penelitian ini hanya terdiri dari satu variabel, yaitu harga. Penentuan variabel independen dan dependen.

3.3.2 Data Training dan Testing

Dalam pembuatan model *Random forest*, langkah selanjutnya adalah membagi data menjadi dua bagian, yaitu data *training* dan data *testing*. Data *training* digunakan untuk melatih model prediksi, sedangkan data *testing* digunakan untuk menguji performa model prediksi tersebut. Umumnya, proporsi data *training* lebih besar daripada data *testing*. Pada penelitian ini, digunakan pembagian proporsi 0,75:0,25, di mana 75% dari data digunakan sebagai data *training* dan 25% sisanya digunakan sebagai data *testing*. Pembagian ini dilakukan secara acak untuk menghindari bias dalam proses pembelajaran dan pengujian model. Didapatkan sejumlah 1389 data *training* dan 463 data *testing* dari total 1852 data, dapat dilihat pada tabel 4.7.

Tabel 2 Hasil Data Training dan Data Testing

Data	Jumlah Data
Data <i>training</i>	1389
Data <i>testing</i>	463

3.3.3 Model *Random forest*

Setelah membagi data menjadi data *training* dan data *testing*, tahap selanjutnya adalah pembuatan model *Random Forest* Dalam penelitian ini, dilakukan pengujian pertama dengan menggunakan beberapa konfigurasi model. Jumlah pohon ($n_estimator$) yang digunakan adalah 50, 100, 150, 200, dan 250. Selain itu, jumlah fitur acak ($random_state$) yang dipilih adalah 42.

Langkah berikutnya adalah mengevaluasi hasil prediksi dari model *Random forest* dengan menggunakan berbagai parameter. Parameter tersebut meliputi jumlah pohon

($n_estimator$) sebanyak 50, 100, 150, 200, dan 250, serta jumlah fitur acak ($random_state$) sebesar 42. Hasil prediksi tersebut kemudian dapat dilihat dalam Tabel 3. Tabel tersebut menampilkan hasil prediksi yang telah diperoleh berdasarkan model *Random forest* dengan parameter-parameter yang telah ditentukan sebelumnya. Tabel ini memberikan gambaran tentang prediksi harga mobil berdasarkan data yang digunakan dan konfigurasi model yang dipilih.

Tabel 3 Hasil Akurasi dan Jumlah tree

Jumlah Tree	Akurasi
250	94,79 %
200	94,77 %
150	94,73 %
100	94,7 %
50	94,56 %

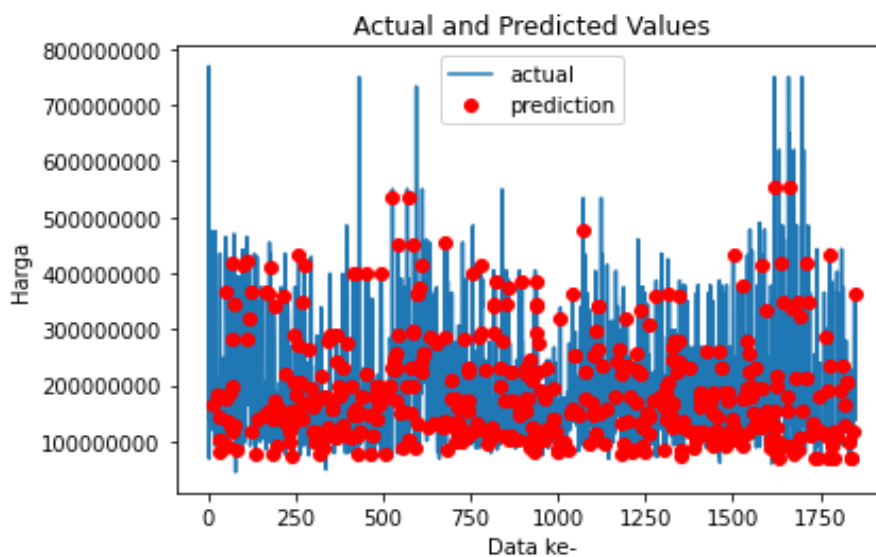
Dalam Tabel 3, diperoleh hasil nilai akurasi terbaik pada jumlah pohon (*tree*) sebanyak 250. Oleh karena itu, dilakukan pelatihan model menggunakan data training dan dilakukan validasi model terhadap data testing. Validasi dilakukan dengan menggunakan parameter default dari *Random forest regressor* pada *library scikit-learn*, yaitu dengan jumlah pohon 250 dan jumlah fitur acak 42.

3.3.4 Evaluasi model

Setelah model prediksi terbentuk selanjutnya melakukan evaluasi model menggunakan metrix. Metrix adalah suatu metode yang digunakan untuk mengevaluasi seberapa baik model prediksi bekerja, adapun metrix yang digunakan dalam penelitian ini adalah :

- AE mengukur selisih absolut antara nilai yang diprediks dan nilai aktual untuk setiap observasi. Metrik ini memberikan gambaran seberapa besar kesalahan prediksi untuk setiap data individu.
- MAE mengukur rata-rata absolut dari selisih antara nilai yang diprediks dan nilai aktual. Metrik ini memberikan gambaran seberapa besar kesalahan prediksi rata-rata yang dihasilkan oleh model.
- MAPE mengukur rata-rata persentase absolut dari kesalahan antara nilai yang diprediks dan nilai aktual. Ini memberikan gambaran seberapa besar kesalahan prediksi rata-rata dalam bentuk persentase relatif terhadap nilai aktual.

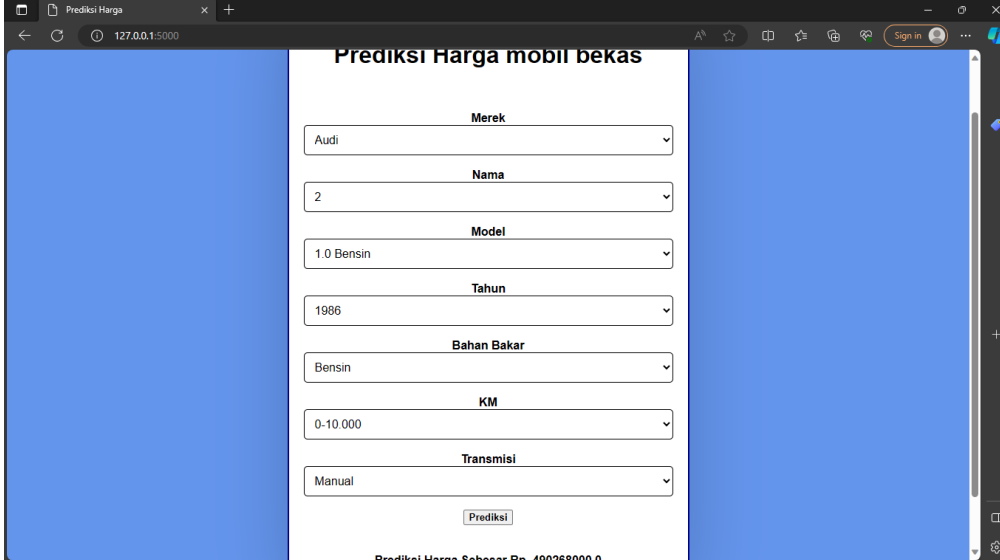
3.3.5 Perbandingan Data Prediksi dan Data Aktual



Gambar 7 Perbandingan Antara Prediksi Data Training Dan Testing Aktual

Grafik pada Gambar 5 menggambarkan perbandingan antara prediksi harga mobil bekas menggunakan data *training* dan data *testing* dengan nilai aktual. Dari grafik tersebut, dapat dilihat bahwa nilai prediksi yang dihasilkan oleh algoritma *Random forest* dengan jumlah pohon (tree) sebanyak 250 hampir mendekati nilai aktual. Meskipun demikian, nilai prediksi yang terbentuk masih sedikit di bawah nilai aktual.

3.3.6 Hasil Tampilan Website



Gambar 8 Halaman From Prediksi

4. Kesimpulan

Berdasarkan hasil analisis dan pengujian pada penelitian ini, dapat diambil beberapa kesimpulan sebagai berikut :

1. Secara keseluruhan, data iklan mobil bekas yang diambil dari website OLX.co.id pada bulan Januari hingga Juni 2023 menunjukkan bahwa mobil dengan merek Toyota merupakan yang paling banyak, dengan jumlah sebanyak 580. Diikuti oleh merek Honda dengan jumlah 348, dan Suzuki dengan jumlah 153. Sisanya terdiri dari berbagai merek lainnya. Dalam hal jenis bahan bakar, mobil dengan bahan bakar bensin mendominasi dengan jumlah 1684, sedangkan mobil dengan bahan bakar solar hanya sejumlah 167. Berdasarkan jenis transmisi, mobil dengan transmisi otomatis paling banyak, dengan jumlah 1462, sedangkan mobil dengan transmisi manual hanya sejumlah 389. Untuk jenis bahan bakar 1684 mobil menggunakan bahan bakar bensin dan 167 mobil menggunakan solar
2. Telah diimplementasikan algoritma *Random forest* untuk memprediksi harga mobil bekas dengan pengujian beberapa konfigurasi model Jumlah Pohon ($n_{estimator}$) Jumlah pohon yang digunakan adalah 50, 100, 150, 200, dan 250 untuk mengukur parameter dalam untuk memprediksi harga mobil bekas yang akurat menggunakan jumlah pohon ($n_{estimator}$) 250 dengan menunjukkan tingkat akurasi sebesar 94,79%. Selain itu, ditemukan Mean Absolute Percentage Error (MAPE) sebesar 5.21% dan Mean Absolute Error (MAE) 9729279.74 yang dimana disimpulkan semakin tinggi parameter yang digunakan maka akan semakin kecil nilai error dalam melakukan pengujian prediksi dan perbandingan antara harga aktual dan harga prediksi dengan selisih rata-rata sebesar 0,10% yang dimana nilai dari hasil prediksi hampir mendekati nilai aktual atau nilai yang sebenarnya.

Daftar Pustaka

- [1] Mohd. S. U. -, R. F. H. -, A. S. -, and S. B. -, "Prediction of Car Purchase based on User Demands using Supervised Machine Learning," *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, vol. 11, no. 1, Jan. 2023, doi: 10.37082/IJIRMP.S.v11.i1.230312.
- [2] A. Nugraha, "Analisis Bauran Pemasaran Dalam Pembelian Mobil Bekas Di Perkasa Mobil," *Jurnal Administrasi Bisnis FISIPOL UNMUL*, vol. 10, no. 4, p. 273, Nov. 2022, doi: 10.54144/jadbis.v10i4.8979.
- [3] B. Widjanarko Otok, J. Statistika, F. Matematika dan Ilmu Pengetahuan Alam, and I. Teknologi Sepuluh November Surabaya Alamat, "Random Forest Dan Multivariate Adaptive Regression Spline (Mars) Binary Response Untuk Klasifikasi Penderita Hiv/Aids Di Surabaya," 2015.
- [4] Agenda Yudha Samudra, "Pendekatan Random Forest untuk model peramalan harga tembakau rajangan di Kabupaten Temanggung," Skripsi, Universitas Sanata Dharma, Yogyakarta, 2019.
- [5] V. Krotov, "Legality and Ethics of Web Scraping," 2018. [Online]. Available: <https://www.researchgate.net/publication/324907302>
- [6] P.-N. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining Instructor's Solution Manual."
- [7] L. Breiman, "Random Forests. Machine Learning," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [8] B. Lareno, "Analisa Dan Perbandingan Akurasi Model Prediksi Rentet Waktu Arus Lalu Lintas Jangka Pendek," *CSRID (Computer Science Research and Its Development Journal)*, vol. 6, no. 3, p. 148, Oct. 2015, doi: 10.22303/csrid.6.3.2014.148-158.
- [9] Sugiyono, *Metode penelitian bisnis*. Jakarta: ALFABETA, 2008.