



Used Car Price Prediction with Random Forest Regressor Model

Bambang Kriswantara¹, Rifki Sadikin^{2*}

Universitas Nusa Mandiri Jakarta dan LIPI Bandung Jawa Barat

Email: 14002426@nusamandiri.ac.id dan rifki.sadikin@gmail.com

Received: February 17, 2022. **Revised:** March 18, 2022. **Accepted:** March 29, 2022. **Issue Period:** Vol.6 No.1 (2022), Pp. 40-49

Abstrac: The pandemic has stopped all activities, including the decline in the level of the world economy, this will reduce people's purchasing power. For those who want to buy a car, a used car can be the second option. In some used car sales showrooms, the main task is to determine price predictions based on historical data during previous transactions. The determinants of car prices are heavily influenced by several attributes in the car, for example: type of fuel, for example, km traveled and so on, this is what causes the price prediction process to take a long time. One of the roles in Macine Learning is being able to learn from previous transaction data and this will be a model that can be used to provide used car price predictions. Car price prediction is included in regression, which is looking for a strong relationship from the influence of variable X (predictor) to variable Y (target). In the prediction, of course, the reality data will not be right with the predicted data, for that in the measurement the model will look for the lowest error rate. The experimental results on the test data using the Random Forest Regressor model resulted in MAE = 1.006 and RMSE = 1.452 while the coefficient of determinant $R^2 = 0.89$. And in the previous study with KNN [18], it produced an error rate of MAE = 2.01 and RMSE = 4.01 and the coefficient of determinant $R^2 = 0.85$. While the comparison model uses Linear Regression, Ridge, Decision Tree and Gradient Boosting. Not all Machine Learning models are suitable for all data, for that it is necessary to choose the right machine learning model by experimenting with several models. And the lowest error level (MAE and RMSE) will be determined. The error values for MAE and RMSE which are close to zero are close to the predicted value close to the actual value. On the other hand, if the error rate is very high, the prediction value is very far from the actual value.

Keywords: Used Car Price Prediction, Machine Learning, Random Forest Regressor

Abstrak: Pandemi membuat semua aktivitas berhenti termasuk menurunnya tingkat perekonomian dunia, ini membuat daya beli masyarakat akan menjadi menurun. Bagi yang ingin membeli mobil, maka mobil bekas bisa mejadi opsi kedua. Pada beberapa showroom penjual mobil bekas yang menjadi tugas utama adalah menentukan prediksi harga berdasarkan data histori selama transaksi sebelumnya. Penentua harga mobil bekas banyak dipengaruhi oleh beberapa atribut dalam mobil tersebut, misalnya: jenis bahan bakar, model transmisi, km tempuh dan sebagainya, inilah yang menyebabkan proses prediksi harga akan menjadi lama. Salah satu peran dalam Macine Learning adalah mampu belajar dari data transaksi sebelumnya dan ini akan menjadi model yang akan dapat digunakan untuk memberikan prediksi harga mobil bekas. Prediksi harga mobil bekas termasuk kedalam



DOI: 10.52362/jisicom.v6i1.752

Works shared below [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

regresi, yaitu mencari hubungan yang kuat dari pengaruh variabel X (predictor) terhadap variabel Y (target). Dalam prediksi tentu tidak akan tepat antara data realitas dengan data prediksi, untuk itu dalam pengukuran model akan mencari tingkat errornya paling rendah. Pada hasil eksperimen pada data test dengan model Random Forest Regressor menghasilkan $MAE=1,006$ dan $RMSE=1,452$ sedangkan koefisien determinan $R^2=0,89$. Dan pada penelitian sebelumnya dengan KNN [18], menghasilkan tingkat error $MAE=2,01$ dan $RMSE=4,01$ dan koefisien determinan $R^2=0,85$. Sedangkan model pembandingan menggunakan Linear Regression, Ridge, Decision Tree dan Gradient Boosting. Tidak semua model Machine Learning cocok terhadap semua data, untuk itu perlu dilakukan pemilihan model machine learning yang tepat dengan cara melakukan eksperimen dengan beberapa model. Dan akan ditentukan tingkat error (MAE dan $RMSE$) yang rendah. Nilai error pada MAE dan $RMSE$ yang mendekati nol berarti nilai prediksi mendekati dengan nilai aktual. Sebaliknya bila tingkat errornya sangat tinggi maka nilai prediksi sangat jauh dengan nilai aktualnya.

Kata Kunci: *Prediksi Harga Mobil Bekas, Machine Learning, Random Forest Regressor*

I. PENDAHULUAN

Paska pandemi corona pasar mobil bekas masih menjadi trend positif, menurut riset permintaan mobil bekas sekitar 15% - 20% (OLX Autos Indonesia, 2021). Kondisi ini tentunya membuat beberapa showroom saling bersaing untuk mendapatkan konsumen. Salah satu hal yang penting dalam jual beli mobil bekas adalah menentukan prediksi harga mobil. Prediksi harga mobil tentu menjadi hal yang penting, dengan mengetahui prediksi maka dapat digunakan untuk mengukur tingkat pendapatan yang akan diterima masa datang, untuk perencanaan dan pengambilan keputusan lain. Data prediksi punya peran sangat penting dalam organisasi dagang, seperti jual beli mobil bekas. Dalam memilih mobil bekas tentunya banyak hal yang menjadi pertimbangan, misalnya harga, merek, tahun produksi, km tempuh, jenis BBM, jenis transmisi, ukuran torsi, kepemilikan dan banyak sekali hal lain yang menjadi dasar pertimbangan untuk menentukan mobil bekas yang akan dibeli. Hal itu tentu akan menentukan harga jual mobil yang ditawarkan. Semakin tinggi kilometer tempuh maka diindikasikan mobil sudah digunakan dengan jarak tempuh yang panjang, dari sisi harga jual tentu akan membuat harga jual akan menurun. Konsumen menganggap mobil dengan kilometer tinggi berarti performanya akan menurun seiring pemakaian. Begitu pula, semakin lama umur kendaraan maka harga jualnya juga semakin menurun.

Dengan banyaknya hal yang dapat mempengaruhi harga mobil bekas, maka prediksi harga mobil bekas akan menjadi hal yang menarik untuk diamati. Bagi showroom mobil bekas, penentuan harga tentunya menjadi hal penting, hal itu dikarenakan banyak hal yang bisa mempengaruhi harga mobil bekas. Disisi lain showroom dituntut bisa bersaing dengan sehat. Masalah yang dihadapi bagi showroom adalah bagaimana bisa dengan cepat menentukan prediksi harga mobil. Bila jumlah mobil bekas yang akan dijual banyak, tentunya akan butuh waktu yang panjang untuk menentukan harga jual mobil. Salah satu masalah yang dihadapi semua showroom adalah menentukan prediksi harga secara cepat dan akurat sehingga showroom bisa menjual mobil dagangannya dan segera mendapatkan revenue. Berbagai showroom saling bersaing harga untuk mendapatkan pelanggan.

Prediksi adalah suatu proses untuk memperkirakan secara sistematis tentang sesuatu yang mungkin terjadi di masa depan berdasarkan informasi data masa lalu [3]. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Pengertian Prediksi sama dengan *forecasting* atau perkiraan. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan dimasa depan dan ini merupakan input bagi proses perencanaan dan pengambilan keputusan.





Untuk melihat apakah sebuah prediksi mempunyai akurasi yang baik dengan melihat tingkat error antara nilai prediksi dengan nilai aktualnya, hal itu dapat dilakukan dengan melakukan evaluasi model dengan MAE yaitu rata-rata nilai absolut error dan RMSE yaitu akar rata2 kuadrat error. Semakin kecil atau mendekati nol maka dapat dikatakan prediksi semakin akurat dan sebaliknya bila nilai errornya semakin besar maka prediksinya adalah jauh dari nilai aktual, sehingga dikatakan buruk.

Machine Learning (ML) merupakan salah satu cabang dari Artificial Intelligence (AI) yang bisa memberikan keputusan berdasarkan data [1] histori. Berdasarkan data histori penjualan mobil bekas dan dengan serangkaian proses permodelan pada machine learning, dengan sejumlah data latih, model ML yang dioptimasi sesuai data tersebut, bisa untuk menghasilkan model prediksi yang baik [1], serta dapat digunakan untuk pengambilan keputusan baik untuk strategi penjualan maupun estimasi revenue yang akan datang. Istilah Machine Learning pertama kali disebutkan oleh Arthur Samuel pada tahun 1959, pada saat itu ia menjelaskan dalam konteks menyelesaikan permainan catur dengan mesin. Secara istilah machine learning merupakan sebuah model komputasi statistik, yang berfokus pada prediksi menggunakan komputer. Algoritma machine learning membangun model matematika dari data sampel, yang dikenal sebagai “data pelatihan atau data training”, untuk membuat prediksi atau keputusan tanpa diprogram secara eksplisit untuk melakukan tugas.

Random Forest Regression/RF merupakan salah satu model regresi yang bisa kita gunakan untuk melakukan prediksi harga mobil. RF merupakan metode klasifikasi berbasis ensemble method dan dibangun dari beberapa decision tree yang berbeda karakteristiknya. Ensemble method sendiri merupakan dimana terdapat beberapa model berbeda yang dilatih untuk memecahkan masalah yang sama dan digabungkan untuk mendapatkan hasil terbaik. Penerapannya pada random forest yaitu metode Random Forest akan membangun beberapa Decision Tree yang dimana setiap membangun 1 tree digunakan observasi dan prediktor yang berbeda dari hasil sampling untuk dilatih dan mencari hasil terbaik. Disamping itu implementasi RF sangat sederhana dan tidak memakan resource (memori yang besar) sehingga lumayan cepat untuk prosesnya sehingga model ini banyak digunakan oleh peneliti untuk pengembangan machine learning.

II. LITERATUR REVIEW

Pada penelitian Ashutosh Datt Sharma dan Vibhor Sharma menggunakan model Multi Linear Regression (MLR), dan dataset dari kaggle dengan jumlah 301 data dan 9 fitur. Dalam eksperimennya tersebut Car_Name di-drop dikarenakan tidak mempunyai pengaruh pada harga jual mobil. Attribute Year diubah menjadi umur pemakaian (Year Age) = Year-2020. Sedangkan attribute Fuel_Type, Seller_Type, Transmission dan Owner dilakukan transformasi dari kategori ke numerik dengan One Hot Encoder. Menurut penelitian tersebut attribute yang digunakan untuk variabel predictor (X) adalah Present_Price, Kms_Driven, Fuel_Type, Seller_Type, Transmission, Owner, Years_Used dan sebagai variabel target (Y) adalah Selling_Price. Pembagian data untuk training dan testing adalah 80% (240) data training dan 20% (61) data testing. Dari hasil eksperimen dengan model Multi Linear Regression didapatkan hasil akurasi 86%[5]. Penelitian berikutnya untuk prediksi harga mobil menggunakan dataset yang sama seperti pada penelitian sebelumnya (vol:02, Nov 2020). Untuk mendapatkan akurasi yang terbaik maka dalam penelitian kali ini, mereka menggunakan komparasi beberapa model regresi seperti, lasso, Ridge, Random Forest, Decision Tree, XGBoost dan Gradient Boosting. Dari eksperimen mendapatkan akurasi tertinggi dengan model Decision Tree 95%[6].

Penelitian Chuancan Chen, Lulu Hao, dan Cong Xu untuk prediksi harga mobil menggunakan 102959 data training dan 19 fitur, dengan metode Linear regression yang menghasilkan tingkat akurasi mencapai 26% sedangkan dengan metode random forest mencapai 52%. Dalam penelitian ini data training dicoba diubah untuk diturunkan menjadi 1250 dan hasil linear regression mencapai 83% sedangkan random forest mencapai 71%.



Disimpulkan dengan bahwa linear regression dan random forest akan lebih baik bekerja pada data yang kecil dan fitur yang kecil [7].

Penelitian Laveena D'Costa, Ashoka Wilson D'Souza, Abhijith K, Deepthi Maria Varghese untuk prediksi harga mobil menggunakan 1870 data dan 7 atribut dengan metode multiple linear regression dan mencapai tingkat akurasi 93.68%, hal ini disebabkan adanya korelasi yang bagus antar feature yang mempengaruhi target(label), seperti Car _Model, Fuel_Type, Emission, Mileage, Ownership, Year_Of_Reg dan Price. Distribusi data yang baik dan merata juga akan mempengaruhi akurasi menjadi lebih baik [8].

Dalam penelitian Mehmet Bilen menggunakan 9925 data dan 7 fitur, dengan model Linear Regression dan ANN (Artificial Neural Network). Tingkat akurasi pada Linear regresi 75% sedangkan ANN 89% untuk prediksi harga mobil[9].

Pada penelitian Muhammad Asgh , Khalid Mehmood , Samina Yasin , Zimal Mehboob Khan menggunakan 25 fitur dan untuk jumlah data tidak disebutkan, dengan model OLS regression dengan tingkat akurasi 89% untuk prediksi harga mobil[10]

Dalam prediksi Nabarun Pal, Dhanasekar Sundararaman harga mobil pada penelitian ini menggunakan 2000 data dan 9 fitur dengan model Random Forest mencapai tingkat akurasi 83,63% [12].

Dalam prediksi harga mobil pada penelitian Prashant Gajera, Akshay Gondaliya, Jenish Kavathiya menggunakan 92386 data, dengan model Random Forest Regressor, Linear Regression, XG Boost Regressor and Decision Tree Regressor. Tingkat akurasi tertinggi pada Decision Tree 93.11% sedangkan XG Boost pada akurasi 92.04%[13].

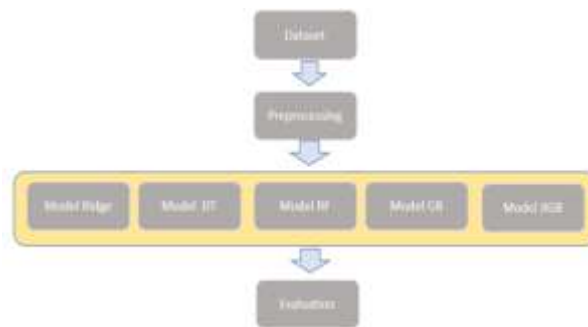
Penelitian Prashant Gajera, Akshay Gondaliya, Jenish Kavathiya prediksi harga mobil menggunakan 92386 data, dengan model Random Forest Regressor, Linear Regression, XG Boost Regressor and Decision Tree Regressor. Tingkat akurasi tertinggi pada Decision Tree 93.11% sedangkan XG Boost pada akurasi 92.04%[13].

Pada penelitian Kandari Puteri, Astried Silvanie dilakukan prediksi harga mobil bekas menggunakan dataset kaggle[18]. Model yang digunakan untuk prediksi harga mobil dengan model KNN regression, dari hasil eksperimen menghasilkan akurasi $R^2 = 0.85$ sedangkan tingkat error RMSE=4,01 dan MAE=2,01.[14]

III. METODOLOGI PENELITIAN

Pada penelitian ini menggunakan eksperimen komparatif, yaitu melakukan eksperimen pengujian terhadap algoritma regresi, kemudian membandingkan beberapa model dengan tingkat akurasi yang terbaik, untuk memprediksi harga mobil bekas. Adapun rancangan atau desain metodologi yang digunakan pada penelitian ini adalah sebagai berikut:





Gambar 1. Rancangan Metodologi Penelitian

A. Dataset

Data bersumber dari kaggle berupa data harga mobil bekas. Jumlah data sebanyak 8128 data dan 13 atribut, yang akan diolah untuk menghasilkan model prediksi harga mobil dengan akurasi yang terbaik. Beberapa atribut yang terdapat pada dataset tersebut adalah:

No	Nama Atribut	Type Data	Keterangan
1	Name	object	nama/jenis mobil
2	Year	int64	tahun pembuatan
3	Selling Price	int64	harga jual
4	Km_driven	int64	jarak tempuh dim KM
5	Fuel	object	jenis bahan bakar (Diesel, Petrol, LPG, CNG)
6	Seller type	object	Dipe penjual (Individual, Dealer, TrustMark)
7	Transmission	object	jenis transmisi (Manual, Automatic)
8	Owner	object	kapemilikan (First, Second, Third, Fourth)
9	Mileage	object	jarak tempuh dalam mil
10	Engine	object	kapasitas silinder
11	Max power	object	maksimal power
12	Torque	object	torsi
13	Seats	float64	jumlah tempat duduk

Table 1. Atribut dan tipe data

B. Data Preprocessing

Pada tahap preprocessing ini, semua atribut akan dilakukan proses agar bisa diproses pada model machine learning. Beberapa proses yang dilakukan pada tahap ini antara lain :

1. Data Transformation

Transformasi digunakan untuk mengubah atribut objek/string/categorical menjadi numerik, machine learning hanya bisa melakukan proses pada data numerik. Oleh sebab itu bila ditemukan atribut yang masih objek/string maka dilakukan proses transformasi ke dalam numerik. Proses ini bisa menggunakan *label encoding* dan *one hot encoder*

a. *Label Encoding* mempunyai teknik mengubah kategori menjadi numerik secara berurutan dari yang bernilai kecil ke besar. Sehingga dalam proses machine learning bobot atau nilai yang besar akan berpengaruh.

b. *Label OneHotEncoding* mempunyai teknik setiap isi kolom akan diubah menjadi atribut baru dan mempunyai nilai 0 dan 1. Angka nol menunjukkan pada atribut tersebut tidak mempunyai nilai sedangkan angka satu menunjukkan bahwa pada atribut tersebut mempunyai nilai.

2. Proses Ekstraksi (pemisahan numerik dan string)

Pada data kategorikal, bila pada atribut berisi numerik dan object, maka perlu dilakukan ekstraksi dengan bantuan *regular expression (regex)* untuk mengambil angka saja. Angka yang telah di ekstrak akan di format menjadi angka float atau integer dan dijadikan atribut baru

3. Data Duplicate

Dilakukan pengecekan pada dataset terhadap data yang bersifat duplikat, data duplicate juga akan mempengaruhi machine learning dalam membuat model pembelajaran dan juga tingkat akurasi yang dihasilkan. Proses untuk melihat data duplicate dalam data frame menggunakan fungsi *duplicate()* dan akan menghasilkan nilai **True** bila terdapat data duplicate dalam dataset tersebut, dan bernilai **False** jika tidak terdapat data duplicate. Untuk menghapus data duplicate bisa menggunakan fungsi *drop_duplicates(inplace=True)*

4. Missing Value

Tujuan pengecekan missing value pada dataset adalah untuk mencari data yang tidak lengkap (*missing value*), karena data yang tidak lengkap akan mempengaruhi dari tingkat akurasi dalam model machine learning. Untuk melihat data yang missing bisa menggunakan fungsi *isnull().sum()*, kemudian bila terdapat data missing maka dilakukan penghapusan dengan fungsi *dropna(inplace = True)*.

5. Fitur Selection

Dalam eksperimen ini menggunakan seleksi fitur untuk memilih fitur yang mempunyai korelasi yang tinggi antara variabel X terhadap variabel Y yang menjadi target. Dengan pemilihan fitur yang terbaik juga akan mempengaruhi dalam proses prediksi.

No	Fitur
1	<i>max_power</i>
2	<i>engine</i>
3	<i>Torque_Nm</i>
4	owner_Test Drive Car
5	transmission_Automatic
6	seller_type_Dealer
7	fuel_LPG
8	owner_Third Owner
9	fuel_CNG
10	fuel_Petrol
11	owner_Fourth & Above Owner
12	fuel_Diesel
13	owner_Second Owner
14	owner_First Owner

Table 2. Fitur yang digunakan

6. Normalisasi

Normalisasi adalah proses penyeimbangan nilai atribut agar proses analisis data lebih akurat. Karena kalau data tidak seimbang artinya ada nilai yang besar maka machine learning akan condong memilih ke angka yang besar. Dengan proses normalisasi ini akan diubah antara 0 sampai dengan 1. Proses normalisasi bisa dilakukan antara lain dengan *min-max*. Cara kerja *min - max* setiap nilai pada sebuah



fitur dikurangi dengan nilai minimum fitur tersebut, kemudian dibagi dengan nilai atau nilai maksimum dikurangi nilai minimum dari fitur tersebut.

7. Cross Validation

Untuk membuat model machine learning maka perlu disiapkan data yang sudah dilakukan proses preprocessing menjadi dua bagian, yaitu bagian training yang berguna untuk melatih pembelajaran algoritma model dan data yang digunakan untuk melakukan testing terhadap model yang sudah dilatih untuk melihat tingkat akurasi prediksi dan melihat tingkat kesalahan yang dihasilkan. *Cross validation* merupakan suatu metode tambahan dari teknik data mining yang bertujuan untuk memperoleh hasil akurasi yang maksimal. Metode ini sering juga disebut dengan k-fold cross validation dimana percobaan sebanyak k kali untuk satu model dengan parameter yang sama. Cross validation prosesnya akan mengulang data test atau data training yang berbeda pada setiap percobaan atau iterasi. Hal ini dimaksudkan agar model bisa dicapai secara seimbang dari setiap percobaan, karena dalam setiap iterasi model sudah mengenal data ujinya.

C. Modeling

Dalam penelitian ini peneliti mengusulkan menggunakan algoritma *Random Forest* untuk pembandingan diantaranya menggunakan *Ridge*, *Decisin Tree*, *Gradient Boosting* dan *XGBoost*. Sehingga akan dipilih algoritma model yang terbaik berdasarkan nilai error yang rendah.

data	supot	remot	nam	engine	seats	mileage	num	max	power	num	torque	ltn	torque	rpm	min	torque	rpm	max
0.2	mpg1015	@m1001	mpg147	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015
0.2	mpg1015	@m1001	mpg147	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015
0.2	mpg1015	@m1001	mpg147	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015
0.2	mpg1015	@m1001	mpg147	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015
0.2	mpg1015	@m1001	mpg147	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015

Gambar 2. Data original

year	selling_price	km_driven	seats	mileage_num	engine_num	max_power_num	torque_ltn	torque_rpm_min	torque_rpm_max
0	2014	45000	14500	5.0	23.40	1248.0	74.00	180.0	2000.0
1	2014	37000	12000	5.0	21.14	1480.0	113.52	250.0	1500.0
2	2006	15800	14000	5.0	17.70	1487.0	78.00	12.7	2700.0
3	2010	22500	12700	5.0	21.00	1360.0	80.00	22.4	1750.0
4	2007	13000	12000	5.0	16.10	1280.0	88.20	11.5	4500.0

Gambar 3. Data setelah transformasi ke numerik



DOI: 10.52362/jisicom.v6i1.752

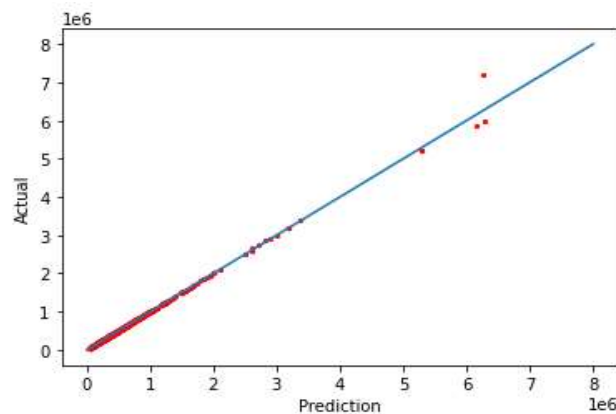
Works shared below [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

IV. Hasil Penelitian

Hasil pengujian dengan model Random Forest pada ekperiment menghasilkan tingkat error MAE=1,006 dan RMSE=1,452. Berikut perbandingan data aktual dengan prediksi:

No	Aktual	Prediksi	Absolut_Error
1	150,000	150,002	0,002
2	325,000	325,000	0
3	160,000	160,000	0
4	200,000	200,000	0
5	949,000	950,000	1
6	825,000	825,000	0
7	200,000	200,000	0
8	175,000	175,000	0
9	300,000	300,000	0
10	1025,000	1026,239	1,239
11	570,000	570,010	0,01
12	290,000	290,000	0
13	399,000	399,080	0,08
14	640,000	640,000	0
15	434,999	434,999	0
16	200,000	200,000	0
17	490,000	490,000	0
18	409,999	409,999	0
19	631,000	630,580	0,42
20	385,000	385,060	0,06

Table 3. Perbandingan data aktual dengan prediksi (Random Forest)



Gambar 4. Grafik antara data aktual dengan prediksi (Random Forest)

V. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan, dapat disimpulkan bahwa machine learning dengan permodelan yang tetap mampu memberikan prediksi regresi terhadap harga mobil bekas. Dataset yang digunakan merupakan data sekunder yang diunduh melalui kaggle.

Pada penelitian ini penulis melakukan eksperimen terhadap lima model regresi (*Ridge*, *Decision Tree*, *Random Forest*, *Gradient Bboosting* dan *XGBOOST*) untuk mencari model yang tepat sesuai data dan bisa memberikan prediksi harga jual mobil bekas dengan akurasi yang baik. Dan model yang paling tinggi akurasinya adalah *Random Forest*

Hasil evaluasi pada model *Random Forest*, Setelah dilakukan pengujian dengan lima model untuk prediksi harga mobil, tingkat error MAE dan RMSE terendah, maka *Random Forest* adalah yang terbaik, tingkat error dari pengukuran MAE=1,006 serta dan RMSE=1,452. Berikut adalah hasil perbandingan tingkat error dengan MAE dan RMSE antar model:

Cross Val	XGBoost		Random Forest	
	MAE	RMSE	MAE	RMSE
1	0,69578232	0,953	0,87533889	1,210
2	0,7538919	1,023	0,58434167	0,770
3	1,05696265	1,412	0,96869167	1,260
4	2,00747218	3,283	1,80665833	3,140
5	0,59614589	1,024	0,79334583	1,390
Average	1,022	1,536	1,006	1,452

Cross Val	Decision Tree		Ridge	
	MAE	RMSE	MAE	RMSE
1	0,66202381	1,020	1,53426244	1,840
2	0,65333333	1,040	1,78009248	2,510
3	1,23170833	1,360	2,46537325	3,400
4	2,65833333	3,590	2,62794185	3,360
5	1,57547980	3,130	1,73196747	2,730
Average	1,356	2,028	2,028	2,768

Cross Val	Gradient Boosting	
	MAE	RMSE
1	1,87533889	1,810
2	0,78434167	0,770
3	0,96869167	1,260
4	1,80665833	3,140
5	0,79334583	1,390
Average	1,246	1,674



REFERENSI

- [1] Jafar Alzubail, Anand Nayyar², Akshi Kumar³"Machine Learning from Theory to Algorithms" doi:10.1088/1742-6596/1142/1/012012, Second National Conference on Computational Intelligence (NCCI 2018)
- [2] Sofianita Mutalib; Nor Aina Azman; Shuzlina Abdul-Rahman "Predicting patients survival using supervised techniques"(DOI: 10.1109/HIS.2011.6122176)Publisher: IEEE,Date of Conference: 5-8 Dec. 2011
- [3] Herdianto. 2013, Prediksi Kerusakan Motor Induksi Menggunakan Metode Jaringan Saraf Tiruan Backpropagation. Medan: Universitas Sumatera Utara.
- [4] A. V. Joshi, Machine Learning and Artificial Intelligence. Cham: Springer International Publishing, 2020.
- [5] Ashutosh Datt Sharma, Vibhor Sharma, "Used Car Price Prediction Using Linear Regression Model", vol:02,Nov 2020, .irjmets).
- [6] Ashutosh Datt Sharma, Vibhor Sharma, Sahil Mittal, Gautam Jain, Sudha Narang* "Predictive Analysis of Used Car Prices Using Machine Learning ", vol:03,June 2021, .irjmets).
- [7] Chuancan Chen, Lulu Hao, Cong Xu, "Comparative Analysis of Used Car Price Evaluation Models", Conference Proceedings 1839, Mey 2017)
- [8] Laveena D'Costa, Ashoka Wilson D'Souza, Abhijith K, Deepthi Maria Varghese, "Predicting True Value of Used Car using Multiple Linear Regression Model", Volume-8, Issue-5S, January 2020
- [9] Mehmet BILEN, "Predicting Used Car Prices with Heuristic Algorithms and Creating a New Dataset",vol:6, 29-43, 2021
- [10] Muhammad Asghar , Khalid Mehmood , Samina Yasin , Zimal Mehboob Khan, "Used Cars Price Prediction using Machine Learning with Optimal Features", Volume: 4, Number: 2, Pages: 113- 119, Year: 2021
- [11] Kanwal Noor, Sadaqat Jan, "Vehicle Price Prediction System using Machine Learning Techniques",Volume 167 – No.9, June 2017
- [12] Nabarun Pal, Dhanasekar Sundararaman, "How much is my car worth? A methodology for predicting used cars prices using Random Forest", Future of Information and Communications Conference (FICC) 2018
- [13] Prashant Gajera, Akshay Gondaliya, Jenish Kavathiya, "OLD CAR PRICE PREDICTION WITH MACHINE LEARNING", Volume:03/Issue:03/March-2021
- [14] K.Samruddhi , Dr. R. Ashok Kumar, "Used Car Price Prediction using K-Nearest Neighbor Based Model", (IJIRASE) Volume 4, Issue 3, DOI: 10.29027/IJIRASE.v4,i3.2020.686-689, September 2020.
- [15] Kandari Puter, Astried Silvanie, "MACHINE LEARNING UNTUK MODEL PREDIKSI HARGA SEMBAKO DENGAN METODE REGRESI LINIER BERGANDA", Vol. 1, No. 2, Oktober 2020
- [16] Siji George C G 1, "Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction" , (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 9, 2020
- [17] Igo Cahya Negara, " Penggunaan Uji Chi-Square untuk mengetahui pengaruh tingkat pendidikan dan Umur terhadap Pengetahuan Penasun mengenai HIV_AIDS di prop DKI",prosiding Seminar Nasional Matematika dan Terapannya 2018 p-ISSN : 2550-0384; e-ISSN : 2550-0392

