

---

**ANALISIS PERFORMANCE MODEL PREDIKSI HARGA JUAL MOBIL BEKAS MENGGUNAKAN MACHINE LEARNING**

Muhammad Ardiansyah Sembiring<sup>1</sup>, Mustika Fitri Larasati Sibuea<sup>2</sup>, Novita elfina<sup>3</sup>  
Universitas Royal

email: <sup>1</sup>adinmantap88@gmail.com, <sup>2</sup>Bukmus.inaction@gmail.com,  
<sup>3</sup>Novie\_elviena@gmail.com

**Abstract:** *Cars are a very popular four-wheeled means of transportation today, so that many consumers or buyers are interested in buying new or used cars depending on their respective economies. One factor that influences consumer interest in buying a car is price. Price greatly influences the sustainability of consumers in buying a car. It is necessary to estimate the estimated price of a used car based on criteria such as mileage, taxes, fuel consumption, and engine capacity. Estimation using regression method where in regression method there are 7 more methods including (1) Linear Regression, (2) Support Vector Regression – Linear, (3) Support Vector Regression – RBF, (4) Decision Tree Regression, (5) Random Forest Regressor, (6) Gradient Boosting Regression, (7) NLP Regressor applied in this research. Based on the 7 regression methods, the best method with the best accuracy value will be sought which will be used in the deployment processing process to determine the price of used cars with a ratio of 90:10, 80:20 and 70:30 producing the best estimated value is decision tree regression. Each method has a high level of accuracy including in the 90:10 ratio decision tree regression as the best method in the ratio has an accuracy level of 99%, and in the 80:20 ratio decision tree regression has an accuracy value of 99%, then the 70:30 ratio decision tree regression again becomes the best method with an accuracy level of 99%.*

**Keyword:** *Machine Learning; Used Car Price Prediction; Regression; Performance Model*

**Abstrak:** Mobil merupakan sebuah alat transportasi kendaraan roda empat yang sangat populer saat ini, sehingga banyak sekali minat konsumen atau pembeli yang ingin membeli mobil baru maupun bekas tergantung dari ekonominya masing-masing. Salah satu yang mempengaruhi minat konsumen dalam membeli mobil yaitu harga. Harga sangat berpengaruh dalam keberlangsungan konsumen dalam membeli suatu mobil. Perlunya upaya estimasi untuk mengetahui perkiraan harga mobil bekas dengan berdasarkan kriteria seperti jarak tempuh, pajak, konsumsi bahan bakar, serta kapasitas mesin. Estimasi menggunakan metode regresi dimana dalam metode regresi terdapat 7 metode lagi meliputi (1) *Linear Regression*, (2) *Support Vector Regression – Linear*, (3) *Support Vector Regression – RBF*, (4) *Decision Tree Regression*, (5) *Random Forest Regressor*, (6) *Gradient Boosting Regression*, (7) *NLP Regressor* yang diterapkan dalam penelitian ini. Berdasarkan 7 metode regresi tersebut akan dicari 1 metode terbaik dengan nilai akurasi paling terbaik yang akan digunakan dalam proses pengolahan deploy untuk menentukan harga mobil bekas dengan rasio yaitu 90:10, 80:20 dan 70:30 menghasilkan nilai estimasi terbaik adalah *decision tree regression*. Masing – masing metode memiliki tingkat akurasi yang tinggi diantaranya dalam rasio 90:10 *decision tree regression* sebagai metode terbaik dalam rasio tersebut memiliki tingkat akurasi sebesar 99%, dan pada rasio 80:20 *decision tree regression* tersebut memiliki nilai akurasi sebesar 99%, selanjutnya rasio 70:30 *decision tree regression* kembali menjadi metode terbaik dengan tingkat akurasi sebesar 99%.

**Kata kunci:** Machine Learning; Prediksi Harga Mobil Bekas; Regresi; Performance Model

## PENDAHULUAN

Diera globalisasi saat ini, banyak sekali manusia membutuhkan sebuah alat transportasi khususnya kendaraan, guna untuk memudahkan manusia dalam melakukan aktivitas. saat ini, sudah banyak teknologi yang diciptakan oleh tangan-tangan manusia. Semua itu dilakukan untuk meringankan segala aktivitas yang dilakukan manusia, agar supaya manusia tidak perlu lagi repot-repot dalam melakukan aktivitas yang sangat melelahkan tujuannya agar terciptanya pekerjaan yang efektif dan efisien serta menghemat waktu. Begitu pesatnya perkembangan teknologi didunia, juga berimbas kepada kemajuan teknologi pada bidang otomotif khususnya kendaraan roda empat yaitu mobil.

Mobil merupakan salah satu alat transportasi yang sebagian besar banyak digunakan pada saat ini. Mobil telah menjadi pilihan utama dalam melakukan kegiatan sehari-hari seperti berpergian jauh. Berdasarkan data Gabungan Industri Kendaraan Bermotor (Gaikindo) penjualan mobil retail pada Februari turun 10,9 persen month to month (mtm) menjadi 69.989 dibandingkan penjualan pada Januari 78.567 persen. Adapun jika dibandingkan dengan tahun lalu penjualan retail naik 49,1 persen penjualan mobil retail pada Februari masih didominasi oleh mobil bekas dengan 21.187 unit, adapun penjualan bulan ini mengalami penurunan 7,5 persen dibandingkan bulan Januari. Penurunan angka yang sangat cukup drastis dan tidak signifikan berimbas kepada sektor perusahaan dibidang otomotif, maka dari itu penerapan teknologi merupakan salah satu solusi terbaik untuk menyokong serta membantu mewujudkan keinginan yang ingin dicapai.

Dalam hal ini, sebagian besar para konsumen dalam memilih mobil khususnya Mobil Bekas terbaik yaitu dengan cara mengikuti trend sekarang dan juga sebagai kebutuhan. Dengan

memperhatikan harga, dan juga pajak kendaraan. sehingga tidak jarang dari para konsumen mengalami dampak yang kurang puas terhadap mobil yang sudah dibeli seperti harga yang overprice dan juga pajak kendaraan yang terlalu mahal dan juga boros bahan bakar.

Dalam pengamatan lokasi yang kami telusuri dilakukan di Air Batu Kab. Asahan, kami mewawancarai salah satu warga yang baru saja membeli mobil Bekas sekitar 1 bulan yang lalu. sama halnya, warga tersebut mengatakan bahwasanya dia membeli mobil karena mengikuti trend zaman sekarang dan juga sebagai kebutuhannya untuk dia bekerja. Namun hal tersebut sangatlah merugikan diri dia sendiri, karena harganya terlalu overprice serta pajak kendaraan nya terlalu mahal sehingga sangat berdampak pada kesulitan ekonominya.

Berangkat dari permasalahan di atas maka peneliti tertarik untuk melakukan proses penelitian guna mendapatkan solusi untuk menentukan harga mobil yang terbaik khususnya mobil bekas yang tidak overprice dan juga sesuai dengan kebutuhan ekonomi. Melalui penerapan teknologi sistem informasi dengan menggunakan Metode Regresi Linier Berganda, dengan melalui penerapan metode ini, peneliti akan memprediksi harga mobil bekas berdasarkan kriteria seperti jarak tempuh, pajak, konsumsi bahan bakar, serta kapasitas mesin. penelitian ini dilakukan untuk menentukan mobil bekas yang dikatakan baik, serta tidak overprice Sehingga akan menghasilkan sebuah informasi tentang harga mobil bekas terbaik berdasarkan semua kriteria tersebut. Melalui pembuatan sistem/aplikasi berbasis web untuk penentuan mobil bekas terbaik serta sesuai dengan kebutuhan ekonomi, maka para konsumen yang nantinya ingin membeli mobil bekas bisa memiliki akses luas untuk memanfaatkan hasil penelitian nantinya.

Regresi Linier Berganda adalah regresi yang memiliki satu variabel dependent (tidak bebas) dan lebih dari satu variabel independent (bebas). Analisis ini untuk mengetahui arah hubungan antara variabel, apakah masing masing variabel independen berhubungan positif atau negatif dan untuk memprediksi nilai dari variabel independen mengalami kenaikan atau penurunan. Data yang digunakan biasanya berskala interval atau rasio (Syahputra et al., 2018).

Regresi linier berganda merupakan suatu analisa yang mempunyai variabel bebas dan lebih dari satu. Teknik regresi linier berganda digunakan dalam mengetahui ada tidaknya dampak signifikan dua atau lebih variabel bebas ( $X_1, X_2, X_3, \dots, X_k$ ) terhadap variabel terikat atau tidak bebas ( $Y$ ) (Siregar, 2021).

Metode penerapan regresi linier berganda dalam bidang otomotif juga telah banyak digunakan salah satunya oleh Rizal Arief Ramadhan(2019) dengan judul “Pengaruh Kualitas Produk, Citra Merek dan Harga Terhadap Keputusan Pembelian Mobil Toyota di Auto 2000 Pasuruan ” Tujuan dari penelitian ini adalah untuk mempengaruhi pengaruh kualitas produk, citra merek dan Harga terhadap keputusan pembelian mobil Toyota. Hasil penelitian ini menunjukkan bahwa variabel kualitas produk, citra merek dan harga berpengaruh positif dan signifikan terhadap keputusan pembelian secara parcial. Sebaiknya perusahaan terus mempertahankan kualitas produknya serta juga memberikan citra merek yang bagus sehingga memberikan kesan baik, agar citra merek dibenak pelanggan terus melekat.

## METODE

Metode riset terdapat beberapa hal penting yang berkaitan pelaksanaan penelitian dan pada tahapan – tahapan yang akan dijalani dalam penelitian ini untuk mendapatkan suatu hasil atau

*output* yang diinginkan. Berikut beberapa hal yang mencakup didalam metode riset antara lain :

### 1. Peralatan Pada Riset

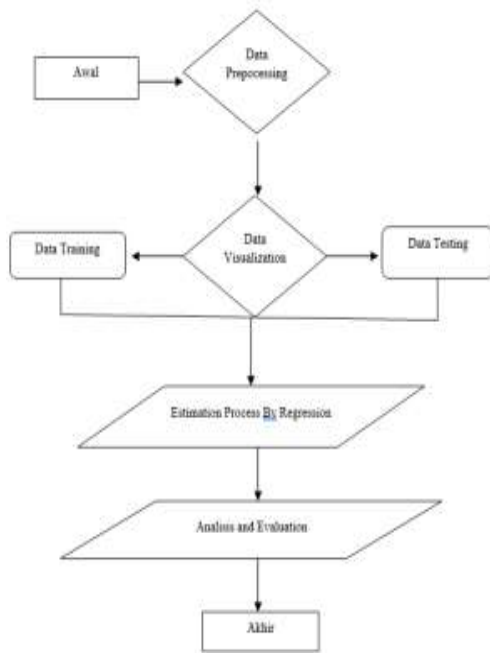
Peralatan yang nantinya akan digunakan pada penelitian ini yaitu seperti Laptop dengan spesifikasi yang mendukung. Adapun *software* yang akan digunakan dalam proses pengolahan data yaitu *software Jupyter notebook* pada *Anaconda Navigator* dimana bahasa pemrograman yang dipakai adalah bahasa pemrograman *Python, Visual Studio Code* sebagai *teks editor* dan menggunakan HTML sebagai bahasa *markup* yang dipakai untuk membangun *deploy* aplikasi tersebut.

### 2. Variabel yang Terdapat Pada Riset

Pada *machine learning*, penelitian ini dikelompokkan sebagai penelitian *supervised learning* yaitu suatu teknik model yang mana data akan diproses memiliki label/target/class untuk mencapai tujuan dalam mengetahui hubungan kausalitas antara variabel bebas (variabel independen) dan variabel yang menjadi target/labelnya (variabel  $y$ ). Maka dari itu, variabel yang terdapat pada penelitian ini terdiri dari variabel independen atau variabel  $x$  yang meliputi *Milage* (jarak tempuh), *Tax* (pajak), *mpg* (konsumsi bahan bakar), dan *Enginesize* (kapasitas mesin). Serta yang menjadi variabel dependennya adalah *Price* (harga).

### 3. Tahapan – Tahapan Riset

Beberapa tahapan yang akan dijalani dalam penelitian ini akan ditampilkan dengan gambaran flowchart sehingga akan terlihat dengan jelas tahapan apa-apa saja yang akan dilakukan.



**Gambar 1. Flowchart Tahapan Sistem Estimasi**

Pada gambar 1, ditampilkan pada gambar flowchart bahwa tahapan awal yang akan dilakukan adalah *Data Preprocessing* yang mana prosesnya terdapat *data cleaning* untuk melakukan *handling missing* pada data, *data selection* yaitu menyeleksi atau mengurangi data berdasarkan kolom yang akan diteliti serta *data transformation*. Tahapan yang kedua adalah melakukan *data visualization* yaitu berfungsi untuk menyajikan representasi pada sebaran data yang digunakan. Tahapan berikutnya adalah melakukan pemisahan pada data set menjadi dua bagian yaitu *data training* dan *data testing*. Selanjutnya, memasuki tahapan proses estimasi dengan memakai regresi yang meliputi tujuh model regresi dan telah melakukan uji coba data dengan menggunakan data testing. Kemudian tahapan terakhir adalah tahapan analisis dan evaluasi data yang menggunakan empat metode evaluasi yaitu MAE, MSE, RMSE dan R2-Score, dengan tujuan untuk mendapatkan hasil akurasi atau akurat pada proses estimasi yang nantinya memperoleh metode terbaik dengan hasil evaluasi yang terbaik pula.

#### 4. Data Riset

Data riset yang nantinya dipakai adalah dataset *mobil\_bekas* yang berasal dari *kaggle.com* pada tahun 2022. Data ini memiliki jumlah sebanyak 1000 baris data. Mempunyai 4 kolom yaitu (*Milage, Tax, Mpg, dan EngineSize*) yang nantinya digunakan pada pengujian dan terdapat 1 kolom (*Price*) sebagai kolom target/label. Sampel dataset yang akan dijabarkan pada penelitian ini menggunakan 5 data teratas yang ditampilkan pada gambar Tabel (1) berikut ini :

**Tabel 1 Sampel Dataset**

mileage	tax	mpg	engineSize	price
24089	265	36.2	2.0	16000
18615	145	36.2	2.0	15995
27469	265	36.2	2.0	13998
14736	150	36.2	2.0	18998
36284	145	36.2	2.0	17498

#### 5. Pembagian Data

Terdapat pembagian data pada penelitian ini dengan menggunakan *data training* dan *data testing*, yang mempunyai rasio bernilai 90:10 yang nantinya *data training* mempunyai data sebanyak 90 dari dataset serta *data testing* sebanyak 10 dari dataset. Algoritma yang memakai rasio 90:10 mempunyai performa yang lebih baik dibandingkan pada algoritma lainnya yang mana tingkat akurasi dapat dilihat sebesar 99,40%. Kemudian rasio 80:20 dan 70:30 yang digunakan pada penelitian ini mempunyai rasio pembagian data lebih bervariasi yang nantinya bertujuan mendapatkan model regresi dengan hasil estimasi terbaik.

#### 6. Estimasi dengan 7 Model Regresi

Model regresi adalah pembantu yang sekarang banyak digunakan dalam mendapatkan hasil akurasi dengan pengaturan sebuah variabel *x* (variabel bebas) dan variabel *y* (variabel terikat). Pada proses estimasi ini memiliki 7 model regresi yang bertujuan sebagai

pembandingan mana model regresi terbaik yang nantinya akan dipergunakan. Adapun 7 model regresi ini adalah :

### 6.1 Linear Regression

*Regresi Linier* yang dibagi menjadi 2 bagian adalah *simple linear regression* dan *multiple linear regression*. Menurut penelitian bahwa regresi linear sederhana adalah model yang menganalisis hubungan antara 1 variabel prediktor dan 1 variabel respon. Dalam penelitian tersebut juga menyatakan bahwa prediksi menggunakan regresi linear sederhana memperoleh hasil prediksi dengan kategori baik karena telah menunjukkan hasil prediksi terbaik selama tahun 2021 pada periode bulan januari. Adapun persamaan pada regresi linear sederhana adalah :

$$Y = a + bx$$

Yaitu :

- Y : variabel dependen
- a : titik perpotongan garis terhadap koordinat y
- b : koefisien variabel independen
- x : variabel independen

Kemudian *multiple linear regression* atau regresi linear berganda merupakan suatu model analisis yang menggambarkan hubungan variabel respon (Y) dengan variabel – variabel yang mempengaruhinya (x) dimana variabel (x) tersebut berjumlah lebih dari satu. Adapun persamaan yang dipakai pada regresi linear berganda sebagai berikut:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$$

Yaitu:

- Y : variabel respon
- X : variabel prediktor yang mempengaruhi variabel respon
- b<sub>0</sub> : intercept
- b : koefisien

### 6.2 Support Vector Regression – Linear

Adalah model regresi yang dipakai dalam memperbaiki *overfitting* dan mempunyai angka akurasi yang baik. *Overfitting* adalah kondisi dimana data saat diolah/*training* mendekati hasil prediksi hampir sempurna. Adapun persamaan yang ada pada model *support vector regression* adalah :

$$f(x) = w^t(p(x)) + b$$

Yaitu :

- f(x) : fungsi regresi
- W<sup>t</sup> : vektor bobot yang memiliki dimensi l
- (p(x)) : titik dalam *space* F, hasil dari pemetaan x pada input *space*
- b : merupakan bias

### 6.3 Support Vector Regression – RBF (Radial Basis Function)

Dalam persamaan pada *support vector regression* yang memiliki fungsi kernel Gaussian-RBF yaitu :

$$K(x_i, x) = \exp\left(-\frac{1}{2a^2} \|x - x_i^T\|^2\right) \quad (2)$$

### 6.4 Decision Tree Regression

Dalam Persamaan pada *decision tree regression* yang pertama mencari nilai entrophy yang ada pada persamaan adalah :

$$mentro(S) = \sum_{i=1} -p(w_i|S) \cdot \log_2(w_i|S)$$

Yaitu :

- S = Himpunan Kasus
- M = Total kelas data
- (w<sub>i</sub>|S) = Proporsi kelas ke-i dalam semua data latihan yang diproses di node S.

Selanjutnya, akan didapatkan angka entrophy dan berikutnya mencari angka gain nya yang tercantum dibawah ini :

$$nGai(S, J) = Entropy(S) - \sum p(v_i|S)$$

$$* E(S_i) \quad i=j$$

Yaitu :

- S : Himpunan kasus
- J : Fitur
- n : Banyak kelas dalam *node* (akar)
- $(v_i|S)$  : Proporsi nilai  $v$  muncul pada kelas dalam *node* (akar)
- $(S_i)$  : Entropi komposisi nilai  $v$  dari kelas ke- $j$  dalam data ke- $i$  node

### 6.5 Random Forest Regressor

Penelitian *random forest* adalah model algoritma yang terdapat dalam regresi dengan tehnik yang menggabungkan prediksi dari beberapa pemecahan yang ada dalam *machine learning* untuk mendapatkan prediksi dengan akurasi terbaik dibandingkan mode tunggal. Kesimpulan dari penelitian tersebut bahwa *random forest regression* lebih akurat dibandingkan pada *linear regression* yang mana rating akurasi pada *random forest* didapat sebesar 97.7% digunakannya perhitungan nilai RMSE dan MAPE. Adapun persamaan *random forest regressor* yang ada adalah :

$$\hat{y}_i = \frac{1}{N_{tree}} \sum_{n=1}^{N_{tree}} \hat{y}_n$$

Yaitu :

- $\hat{y}_i$  = hasil prediksi
- $N_{tree}$  = total jumlah pohon
- $\hat{y}_n$  = hasil prediksi pohon ke- $n$

### 6.6 Gradient Boosting Regression

Adalah salah satu dari algoritma ensemble yang menggunakan peningkatan akurasi dengan sebuah nilai. Model *gradient boosting regression* dapat mengatasi pola yang kompleks atau rumit. Penyusunan data dari *gradient boosting regression* yaitu pohon keputusan (*decision tree*). Kemudian persamaan pada *gradient boosting regression* ini meliputi :

$$r_{im} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

for  $i=1, \dots, n$

### 6.7 Natural Language Processing (NLP) Regressor

Merupakan kelebihan dalam program komputer yang bertujuan memproses bahasa pada pengguna dalam lisan atau tulisan yang mendesain atau membuat suatu aplikasi yang nantinya dipergunakan untuk melengkapi komunikasi pengguna dengan mesin melalui bahasa alami yang sering dipakai pengguna.

### 7. Model Pengukuran Akurasi Menggunakan Nilai Error

Dataset yang sudah diolah dengan 7 model regresi, selanjutnya akan diukur nilai akurasinya. Data yang diolah bersifat estimasi yang terkandung dalam unsur ketidaktepatan maka dibutuhkan untuk memasukkan nilai *error* sebagai tolak ukur akurasi pada estimasi. Nilai *Error* merupakan perselisihan antara nilai pengamatan yang sesungguhnya dengan nilai prediksi. Dalam estimasi yang diukur memakai MAE, MSE, dan RMSE yang nilainya sangat baol adalah nilai yang paling kecil, semakin kecil nilai yang didapatkan maka hasil estimasi lebih bagus. Lain dengan *R2-Score* yaitu nilai koefisien determinan yang mendekati 1 dapat diartikan bahwa variabel independennya memberikan seluruh informasi yang dibutuhkan untuk mengestimasi variabel dependennya. Nilai *error* yang nantinya dipergunakan adalah 4 (empat) model pengukuran akurasi nilai estimasi sebagai berikut :

#### 7.1 Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - F_i|$$

Yaitu :

- $n$  : ukuran sampel
- $A_i$  : nilai data aktual ke- $i$
- $F_i$  : nilai data estimasi ke- $i$

#### 7.2 Mean Square Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Yaitu :

$n$  : jumlah data

$Y_i$  : nilai yang diamati

$\hat{Y}_i$  : nilai estimasi

### 7.3 Root Mean Square Error (RMSE)

$$RMSE = \left( \frac{\sum (y_i - \hat{y}_i)^2}{n} \right)^{1/2}$$

Yaitu :

$y_i$  : nilai hasil pengamatan

$\hat{y}_i$  : nilai hasil estimasi

$N$  : jumlah data

### 7.4 R<sup>2</sup>-Score

$$R^2 = 1 - \frac{RSS}{TSS}$$

Yaitu :

$R^2$  : koefisien determinan

RSS : jumlah kuadrat

TSS : total kuadrat

## HASIL DAN PEMBAHASAN

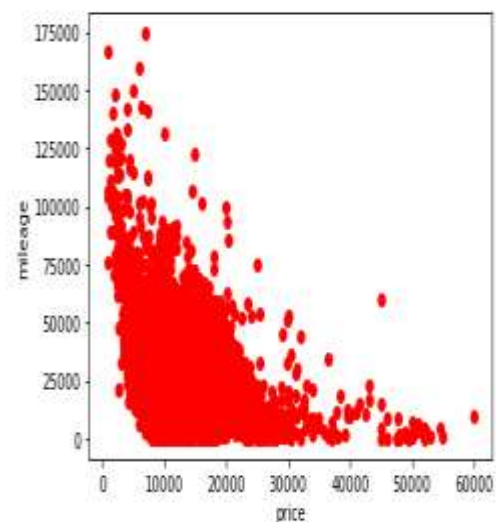
Dalam pembahasan penelitian ini akan dikerjakan visualisasi data yang nantinya data akan terasa sangat mudah untuk diamati serta dapat memahami hubungan antara tiap variabel x terhadap variabel y. Kemudian, data akan diuji menggunakan 4 (empat) bentuk penguraian data dengan rasio yang berlainan satu dengan yang lain dan dimana didalamnya akan dilakukan pengujian data menggunakan 7 model regresi serta diukur dengan 4 model nilai evaluasi. Hal ini mempunyai tujuan untuk menemukan model regresi yang mana diantaranya paling baik dengan menghasilkan nilai akurasi yang lebih terakurat untuk digunakan dalam price (harga) pada kendaraan mobil bekas. Beberapa tahapan pembahasan yang dilakukan yaitu :

### 1. Visualisasi Data

Visualisasi data memiliki tujuan untuk mempermudah seseorang dalam memantau dan mempelajari sebuah data. Ada berbagai model yang dapat digunakan untuk menyajikan suatu data kedalam bentuk visualnya. *Scatter plot* menggunakan *library* yaitu “*matplotlib.pyplot*”. Dalam penelitian ini visualisasi data menggunakan *scatter plot* dengan tujuan agar dapat melihat plot sebaran data dan meninjau seberapa besar hubungan pengaruh antara sumbu x dengan sumbu y. Data *mobil\_bekas* yang digunakan untuk mengestimasi harga mobil bekas yang telah divisualisasikan terhadap sumbu x nya. Visualisasi data disajikan dalam bentuk antar variabel x dengan variabel y dan penyajian *scatter* tiap variabel menggunakan warna yang berbeda agar lebih mudah dipahami dan mudah diingat. Visualisasinya data seperti dibawah ini:

#### 1.1 Scatter Plot Milage Terhadap Price

Seperti berikut ini adalah tampilan visualisasi data pada *milage* sebagai sumbu y terhadap *Price* sebagai sumbu x.



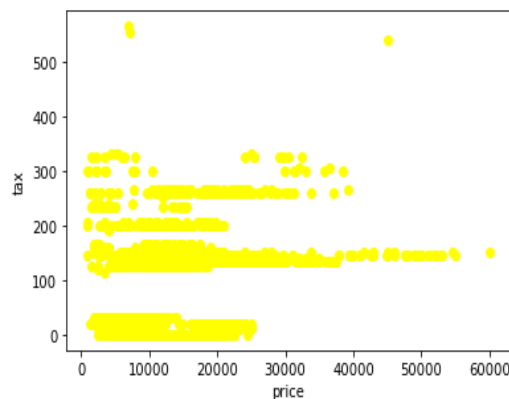
**Gambar 2 Scatter Milage terhadap Price**

Penjelasan gambar diatas, plot titik sebaran *milage* (sumbu y) divisualisasikan dengan titik berwarna merah. Berdasarkan gambar data diatas, dapat dilihat bahwa ukuran jarak tempuh semakin besar pada sebuah kendaraan mobil bekas maka semakin berkurang

juga harga jualnya, dikarenakan menyangkut kondisi mesin yang terus-terusan beroperasi. Hubungan yang ditampilkan oleh gambar diatas adalah hubungan negatif karena adanya penurunan yang ditunjukkan oleh sumbu x (*price*) yang diikuti dengan peningkatan sumbu y (*milage*).

### 1.2 Scatter Plot Tax Terhadap Price

Kemudian tampilan visualisasi data pada *tax* (sumbu y) terhadap *price* (sumbu x)

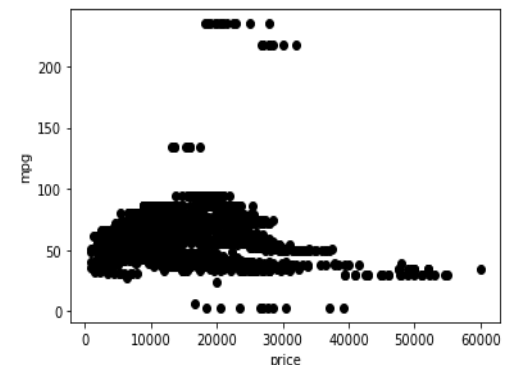


Gambar 3 Scatter Tax terhadap Price

Visualisasi plot titik sebaran dari sumbu y (*tax*) terhadap sumbu x (*price*) disajikan kedalam *scatter* berwarna kuning. Hubungan yang ditunjukkan bersifat positif karena masih menunjukkan peningkatan baik dari variabel *tax* (pajak) maupun variabel *price*.

### 1.3 Scatter Plot Mpg Terhadap Price

Tampilan visualisasi data pada *mpg* (sumbu y) terhadap *price* (sumbu x)



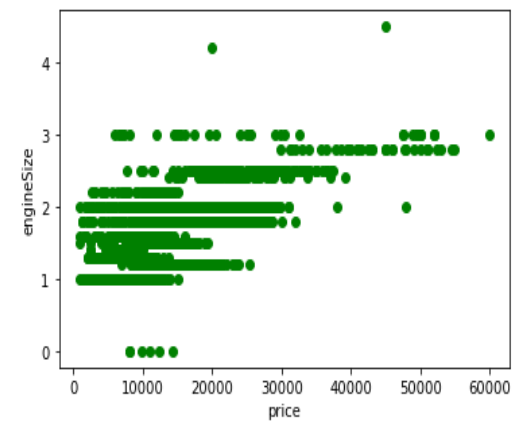
Gambar 4 Scatter Mpg terhadap Price

Visualisasi pada sumbu y (*mpg*) terhadap sumbu x (*price*) menggunakan

*scatter* plot berwarna black. Hubungan yang ditampilkan bersifat positif karena masih menunjukkan baik dari variabel (konsumsi bahan bakar) maupun variabel *price*.

### 1.4 Scatter Plot EngineSize Terhadap Price

Tampilan visualisasi data pada *enginesize* (sumbu y) terhadap *price* (sumbu x)



Gambar 4 Scatter EngineSize terhadap Price

Visualisasi pada sumbu y (*enginesize*) terhadap sumbu x (*price*) menggunakan *scatter* plot berwarna green. Hubungan yang ditampilkan bersifat positif karena masih menunjukkan baik dari variabel (kapasitas mesin) maupun variabel *price*.

## 2. Pengujian Data dengan 7 Model Regresi

Dalam penelitian ini menggunakan metode regresi yang mana memperbandingkan 7 metode regresi dalam memperoleh model hasil akurasi yang paling terbaik untuk digunakan dalam proses prediksi harga jual dari kendaraan mobil bekas. Penelitian ini memakai pembagian data, yang akan diuji sebanyak 4 bentuk seperti pembagian rasio awal yaitu 90:10, kedua 80:20, dan ketiga 70:30. Data yang teruji dengan rasio yang bervariasi untuk mendapatkan serta menentukan model regresi yang mana paling sering muncul dalam

menghasilkan nilai estimasi yang terbaik. Sehingga, nantinya akan diketahui model regresi yang terbaik dalam melakukan estimasi terhadap dataset untuk mencari harga jual mobil bekas. Berikut pembahasan penelitian untuk menemukan model regresi terbaik yang diuraikan kedalam rincian poin sebagai berikut :

## 2.1 Pengujian Akurasi Pada Rasio Data 90:10

Pengujian data awal menggunakan rasio 90:10, dimana *data training* sebesar 90% dari total data dan *data testing* sebesar 10% menggunakan 7 model regresi diperoleh hasil seperti pada tabel dibawah ini :

**Tabel 2 Pengujian Akurasi Data Rasio 90:10**

PEN GU KU RA N	MODEL						
	LinearRe gression	SVR- Linear	SVR- RBF	Decisio nTree	Random Forest	GradientB oosting	MLP- Regressor
MA E	2.85E+0 3	3.86E+0 3	5.34E+0 3	8.21E- 01	5.78E+0 2	1.01E+03	5.64E+03
MS E	1.17E+0 7	2.75E+0 7	4.19E+0 7	5.25E+0 1	6.74E+0 5	1.73E+06	4.45E+07
RM SE	3426.092 644	5245.80 2766	6474.67 5	7.24657 302	821.099 9967	1316.703 487	6669.9578 33
R2- SCO RE	0.719290 41	0.34191 352	- 0.00252 45	0.99999 874	0.98387 68	0.958539 45	- 0.0639108 3

Ketentuan yang digunakan untuk mengetahui hasil estimasi yang baik adalah dalam nilai pengukuran menggunakan MAE, MSE, RMSE yaitu nilai paling kecil yang dihasilkan merupakan nilai dengan hasil estimasi yang paling bagus. Berbeda dengan R2-Score yaitu nilai yang paling mendekati angka 1 (satu) dan jauh dari angka 0 (nol) merupakan nilai dengan estimasi yang paling bagus. Pada gambar tabel diatas, model pengukuran MAE nilai estimasi yang paling bagus ditunjukkan oleh metode *Gradient Boosting Regressor* dengan nilai estimasi sebesar 1.01E+03 dimana nilai ini merupakan nilai terkecil pada model pengukuran MAE. Kemudian, untuk model pengukuran MSE nilai estimasi yang paling bagus ditunjukkan oleh metode *Linier Regressio* dengan nilai estimasi sebesar 1.17E+07 dimana nilai ini merupakan nilai terkecil pada model pengukuran MSE. RMSE metode yang menghasilkan nilai estimasi yang paling bagus ditunjukkan oleh metode *Decision Tree Regressor* dengan nilai estimasi sebesar 7.24657302 dimana nilai ini

merupakan nilai terkecil pada model pengukuran RMSE. Selanjutnya, model pengukuran dengan R2-Score nilai estimasi yang paling mendekati angka 1 (satu) masih dihasilkan oleh perhitungan dengan metode *Decision Tree Regressor* dengan hasil estimasi sebesar 0.99999874. Berdasarkan pada pengamatan dan penampilan gambar tabel diatas, pengujian akurasi dengan rasio data 90:10 menghasilkan *Decision Tree Regressor* sebagai metode regresi yang lebih baik karena metode ini paling banyak menghasilkan nilai estimasi terbaik didalam 2 model pengukuran yaitu RMSE dan R2-Score. Sedangkan terbaik kedua dalam pengujian rasio 90:10 ini adalah *Gradient Boosting Regressor* dan *Linier Regressio* dengan masing-masing menghasilkan 1 nilai estimasi terbaik pada model pengukuran MAE dan MSE.

## 2.2 Pengujian Akurasi Pada Rasio Data 80:20

Pengujian data kedua menggunakan pembagian data dengan

rasio 80:20. Terdapat sebanyak 80% *data training* dan 20% *data testing* dari dataset, diperoleh hasil seperti tabel dibawah ini :

**Tabel 3 Pengujian Akurasi Data Rasio 80:20**

PEN GU KU RAN	MODEL						
	LinearR egressio n	SVR- Linear	SVR- RBF	Decisio nTree	Random Forest	GradientB oosting	MLP- Regressor
MA E	2849.72 52	4310.611 935	5658.47 97	9.09935 897	581.455 7692	1016.437 946	5619.861 002
MSE	1.45E+ 06	3.19E+07	4.57E+0 7	5.94E+0 3	8.22E+0 5	1.65E+06	4.79E+07
RMS E	3383.47 9413	5651.183 124	6763.57 25	77.0929 0308	906.652 6091	1284.889 165	6924.381 011
R2- SCO RE	0.75245 618	0.309436 26	0.01081 55	0.99987 149	0.98222 511	0.964300 98	- 0.036780 75

Pada gambar tabel diatas, didapatkan metode *Decision Tree Regression* sebagai metode regresi terbaik pada 7 metode regresi lainnya. Hal tersebut dapat kita lihat bahwa nilai estimasi *Decision Tree Regression* yang lebih baik pada 3 model pengukuran yaitu nilai sebesar 9.09935897 pada MAE, 77.09290308 pada RMSE, dan 0.99987149 pada R2-SCORE yang mana dua nilai diantaranya merupakan nilai yang paling kecil dan satu nilai yang lebih mendekati angka 1 dan berdasarkan hasil daripada metode regresi lainnya serta dalam pengukuran menggunakan *MSE Linier Regressio* yaitu bernilai sebesar 1.45E+06 yang menjadi

metode terbaik kedua dalam pengujian tersebut. Maka, didapat kita lihat dalam pengujian akurasi data 80:20 dihasilkannya metode *Decision Tree Regression* sebagai metode dengan nilai akurasi yang paling bagus.

### 2.3 Pengujian Akurasi Pada Rasio Data 70:30

Pengujian data ketiga dengan menggunakan rasio data 70:30 yaitu 70% *data training* dan 30% *data testing* menggunakan 7 model regresi diperoleh hasil sebagai berikut :

**Tabel 4 Pengujian Akurasi Data Rasio 70:30**

PE NG UK UR AN	MODEL						
	LinearRe gression	SVR- Linear	SVR- RBF	Decisio nTree	Random Forest	Gradient Boostin g	MLP- Regressor
MA E	2599.357 198	4081.818 763	5408.11 73	6.06623 932	483.646 1539	1015.72 0829	5023.812 698
MS E	8.71E+05	2.91E+07	4.30E+0 7	3.96E+0 3	5.36E+0 5	1.63E+0 6	4.05E+07
RM SE	3141.884 003	5391.281 24	6554.23 89	62.9460 9178	731.795 4041	1274.97 5651	6367.524 63
R2- SC OR E	0.772260 47	0.329433 12	0.00893 38	0.99990 859	0.98764 515	0.96249 735	0.064595 64

Pada gambar tabel diatas, model pengukuran MAE nilai estimasi yang paling bagus ditunjukkan oleh metode *Decision Tree Regressor* dengan nilai estimasi sebesar 6.06623932 dimana nilai ini merupakan nilai terkecil pada model pengukuran MAE. Kemudian, untuk model pengukuran MSE nilai estimasi yang paling bagus ditunjukkan oleh metode *Gradient Boosting Regressor* dengan nilai estimasi sebesar 1.63E+06 dimana nilai ini merupakan nilai terkecil pada model pengukuran MSE. RMSE dan R2-Score metode yang menghasilkan nilai estimasi yang paling bagus ditunjukkan oleh metode *Decision Tree Regressor* dengan nilai estimasi sebesar 62.9460918 serta dengan nilai yang paling mendekati angka 1 (satu) dengan hasil estimasi sebesar 0.99990859.

### 3. Hasil Metode Regresi Terbaik

Setelah melakukan pengujian akurasi dengan 3 bentuk rasio data, kemudian diperoleh metode terbaik yang paling banyak menghasilkan nilai estimasi terbaik adalah *decision tree regression*. Masing – masing metode memiliki tingkat akurasi yang tinggi diantaranya dalam rasio 90:10 *decision tree regression* sebagai metode terbaik dalam rasio tersebut memiliki tingkat akurasi sebesar 99%, dan pada rasio 80:20 *decision tree regression* tersebut memiliki nilai akurasi sebesar 99%, selanjutnya rasio 70:30 *decision tree regression* kembali menjadi metode terbaik dengan tingkat akurasi sebesar 99%. Berikut ini hasil nilai terbaik dengan tingkat akurasi yang dihasilkan :

**Tabel 5 Hasil Estimasi Metode Terbaik**

Metode Regresi	90:10	80:20	70:30
<i>Decision Tree Regression</i>	99%	99%	99%

## SIMPULAN

Mobil merupakan sebuah alat transportasi kendaraan roda empat yang sangat populer saat ini, sehingga banyak sekali minat konsumen atau pembeli yang ingin membeli mobil baru maupun bekas tergantung dari ekonominya masing-masing. Salah satu yang mempengaruhi minat konsumen dalam membeli mobil yaitu harga. Harga sangat berpengaruh dalam keberlangsungan konsumen dalam membeli suatu mobil. Untuk itu penelitian ini bertujuan untuk memperoleh hasil estimasi untuk menentukan harga jual mobil bekas dengan menggunakan 7 metode regresi. Variabel yang digunakan terdiri dari variabel independen meliputi *Jarak Tempuh, Pajak, Konsumsi Bahan Bakar, serta Kapasitas Mesin* dan variabel dependen yang menjadi labelnya

adalah *Harga*. Estimasi menggunakan 7 metode regresi untuk memperoleh metode manakah yang menghasilkan nilai akurasi yang paling akurat.

Berdasarkan hasil pengolahan data menggunakan 7 metode regresi menggunakan *software jupyter notebook, visual studio code* dan *python* sebagai bahasa pemrogramannya serta pengujiannya menggunakan 4 rasio pengukuran nilai akurasi yaitu 90:10, 80:20 dan 70:30 maka diperoleh metode terbaik yang menghasilkan nilai akurasi paling bagus yaitu metode *Decision tree regression* Terpilihnya menjadi metode terbaik dikarenakan metode tersebut paling sering muncul dalam memperoleh nilai estimasi terbaik pada pengujian dengan rasio 90:10, 80:20 dan 70:30. Hasilnya metode *Decision tree regression* keluar sebagai metode terbaik karena metode

tersebut paling banyak menghasilkan nilai akurasi terbaik pada saat pengujian akurasi dengan menggunakan 3 rasio. *Decision tree regression* pada pengujian akurasi rasio 90:10 menghasilkan 2 nilai akurasi terbaik yaitu nilai estimasi pada RMSE sebesar 7.24657302 serta pada R2-Score hasil estimasi sebesar 0.99999874 atau tingkat akurasinya sebesar 99%. Kemudian pada rasio 80:20 *Decision tree regression* menghasilkan 3 nilai akurasi terbaik yaitu nilai estimasi pada MAE sebesar 9.09935897 dan estimasi pada RMSE sebesar 77.09290308 serta pada R2-Score hasil estimasi sebesar 0.99987149 atau tingkat akurasinya sebesar 99%. Kemudian pada rasio 70:30 *Decision tree regression* menghasilkan 3 nilai akurasi terbaik yaitu nilai estimasi pada MAE sebesar 6.06623932 dan estimasi pada RMSE sebesar 62.9460918 serta pada R2-Score hasil estimasi sebesar 0.99990859 atau tingkat akurasinya sebesar 99%. Dengan demikian metode regresi yaitu *Decision tree regression* digunakan sebagai metode regresi yang akan di implementasikan dalam pembuatan *deploy* aplikasi untuk mengestimasi harga jual mobil bekas.

#### DAFTAR PUSTAKA

- Siregar, A. Z. (2021). Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru. ... : *Jurnal Penerapan Sistem Informasi (Komputer Dan ...*, 2(3), 133–137. <http://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/73>
- Syahputra, T., Halim, J., & Perangin-angin, K. (2018). Penerapan Data Mining Dalam Memprediksi Tingkat Kelulusan Uji Kompetensi ( UKOM ) Bidan Pada STIKes Senior Medan Dengan Menggunakan Metode Regresi Linier Berganda. *Sains Dan Komputer (SAINTIKOM)*, 17(1), 1–7.
- Rivandi, Ahmad, Efori Bu'ulolo, and Natalia Silalahi. "Penerapan Metode Regresi Linier Berganda Dalam Estimasi Biaya Pencetakan Spanduk (Studi Kasus: PT. Hansindo Setiapratama)." *Pelita Informatika: Informasi dan Informatika* 7.3 (2019): 263-268.
- Mulyani, Evi Dewi Sri, et al. "Estimasi Harga Jual Mobil Bekas Menggunakan Metode Regresi Linier Berganda." *E-JURNAL JUSITI: Jurnal Sistem Informasi dan Teknologi Informasi* 9.1 (2020): 1-8.
- Tsani, Anindya Aulia Rizqy, and Firman Farid Muhsoni. "Estimasi stok karbon mangrove di desa Taddan Kecamatan Camplong Kabupaten Sampang." *Jurnal Ilmu Kelautan Kepulauan* 5.1 (2022).
- NUHA, HILAL HUDAN, and AULIA ARIF WARDANA. "Estimasi Utilisasi Prosesor pada Jaringan Interkoneksi Optik menggunakan Regresi Gaussian." *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika* 10.3 (2022): 702.
- Boy, Ahmad Fitri. "Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara)." *Journal of Science and Social Research* 3.2 (2020): 78-85.
- Nafi'iyah, Nur. "Prediksi jumlah

penjualan pada toko makmur jaya elektronik dengan regresi linier."

RESEARCH: Journal of Computer, Information System & Technology Management 2.2 (2019): 47-50.

Bakti, Wisnu Nugroho Setya.  
Komparasi Algoritma Regresi

Linier Berganda Dan Decision Tree Regressor Untuk Memprediksi Harga Mobil Bekas. Diss. Universitas Mercu Buana Jakarta, 2022.