

Perbandingan Kecocokan Regresi terhadap Tahun Produksi dan Harga Mobil

Reyza Farzan Rahmatsyah

Statistics Department

BINUS UNIVERSITY

DKI Jakarta, 11530

reyza.rahmatsyah@binus.ac.id

Sherryl Kurniawan

Statistics Department

BINUS UNIVERSITY

DKI Jakarta, 11530

sherryl.kurniawan@binus.ac.id

Rasyad Muhammad Ramdhanazuri

Statistics Department

BINUS UNIVERSITY

DKI Jakarta, 11530

rasyad.ramdhanazuri@binus.ac.id

Dira Abiyyu Bagaspati

Statistics Department

BINUS UNIVERSITY

DKI Jakarta, 11530

Dira.bagaspati@binus.ac.id

Felix Indra Kurniadi, S.Kom., M.Kom

Scientific Department

BINUS UNIVERSITY

DKI Jakarta, 11530

felix.indra@binus.ac.id

Abstract— The price of cars has been steadily increasing since its inception in 1886. This paper examines the price of cars in Africa's fastest growing economic country, Nigeria. There are three goals in mind in this paper. First, to investigate the relationship between the manufacturing year of the car relative to its price. Second, to find the advantages and disadvantages of different regression methods. And thirdly, to find the Regression model that best fits the data. While incorporating the process of data mining, this paper found that there was a positive relation between the year of manufacture and the price of the car. With this in mind, we inserted this data into three regression models, the linear regression model, the loess regression, and the robust Regression. It found that the linear model has limitations in predicting the price of the car. It found that the Loess method is the most ideal to predict the outcome. This is because the loess method can produce an accurate result with visible outliers.

Abstract - Harga mobil meningkat sejak penemuannya pada tahun 1886. Makalah ini mengkaji harga mobil di negara dengan pertumbuhan ekonomi tercepat di Afrika, Nigeria. Menggunakan data ini, ada tiga tujuan dalam pikiran. Pertama, untuk menyelidiki hubungan antara tahun pembuatan mobil relatif terhadap harganya. Kedua, untuk mengetahui kelebihan dan kekurangan metode regresi yang berbeda. Dan ketiga, untuk menemukan model Regresi yang paling sesuai dengan data. Dengan menggabungkan proses penambangan data, makalah ini menemukan bahwa terdapat hubungan positif antara tahun pembuatan dan harga mobil. Dengan mempertimbangkan ini, kami memasukkan data ini ke dalam tiga model regresi, model regresi linier, regresi loess, dan metode kuat untuk Regresi linier. Ditemukan bahwa model linier memiliki keterbatasan dalam memprediksi harga mobil. Ditemukan bahwa metode Loess adalah yang paling ideal untuk memprediksi hasil. Hal ini dikarenakan metode loess dapat memberikan hasil yang akurat dengan outlier yang terlihat.

Keywords— Cars, Linear Regression, Data Mining. Price

I. INTRODUCTION

Perusahaan Otomotif menjadi jembatan penghubung bagi individu dalam membeli dan menjual sebuah kendaraan. Dalam sektor otomotif, menganalisis harga pasar memegang peranan yang penting bagi perusahaan dan individu yang berkontribusi di dalam industri otomotif. Dengan adanya analisis mengenai harga produk tersebut, diharapkan dapat menjadi pembanding terhadap harga pasar sebelum membeli ataupun menjual mobil bekas. Dengan menggunakan beberapa pendekatan model linear regresi, analisis sebuah perusahaan mobil terhadap harganya dapat dilakukan. Berdasarkan dari data, nominal harga dari sebuah produk mempengaruhi keputusan konsumen dalam menentukan produk yang dipilih. Selain harga produk, kualitas produk tersebut juga mempengaruhi keputusan konsumen dalam menentukan produk yang ingin dibeli.

Menggunakan Uji Asumsi Klasik linear model dan outlier test untuk menyelesaikan analisis ini. Untuk Uji Asumsi Klasik, memanfaatkan regresi linear untuk variabel numerical, dengan 4 tipe uji, yaitu uji normalitas residual, uji autokorelasi, uji multikolinearitas, dan uji homogenitas, dan untuk Outlier test digunakan untuk mengobservasi sebuah data yang unik dan karakteristik yang berbeda yang ditemukan di dalam dataset.

Tujuan pengujian asumsi klasik ini adalah untuk memberikan kepastian bahwa persamaan regresi yang didapatkan memiliki ketepatan dalam estimasi, tidak bias dan konsisten. Asumsi klasik adalah syarat-syarat yang harus dipenuhi pada model regresi linear OLS agar model tersebut menjadi valid sebagai alat penduga.

Maka dalam data *Nigerian_price_car* terdapat data yang berisi tentang aspek aspek penjualan kendaraan yang meliputi merk mobil, tahun pembuatan, kondisi, jarak tempuh, kapasitas mesin, harga, transmisi, bahan bakar yang dimana

pada data tersebut langkah pertama yaitu membersihkan data dari Outlier karena data yang di terima masih dalam data mentah kemudian memberikan visualisasi data yang bertujuan untuk melihat keterlibatan 1 data dengan data yang lain.

Setelah itu melakukan uji asumsi pada data, hipotesis data, linear regression models, Measurements. Yang bertujuan untuk menghasilkan data yang sesuai yang kita inginkan.

II. PREVIOUS RESEARCH

Harga merupakan variabel yang memegang peranan penting dalam pemasaran, harga dapat mempengaruhi pelanggan dalam mengambil setiap keputusan untuk membeli produk, salah satunya yaitu mobil. Untuk melihat besarnya pengaruh harga terhadap keputusan pelanggan dalam melakukan pembelian mobil, beberapa perusahaan telah melakukan riset dan analisis untuk membuktikan pengaruh harga terhadap keputusan pelanggan.

Riset yang pertama [3], dilakukan oleh perusahaan mobil Nissan, dimana menggunakan linear regresi untuk membuktikan keputusan pembelian mobil nissan pada PT. Wahana Wirawan Manado dengan populasi 110 responden dan sampel yang digunakan sebanyak 55 responden. Untuk melihat pengaruhnya, perusahaan ini melihat koefisien korelasi (R) sebagai acuannya, dengan mencari pola hubungan dua variabel atau lebih. Dengan menggunakan 4 indikator, yaitu harga yang terjangkau, mampu bersaing dengan perusahaan lain, faktor pendukung pengambilan keputusan, dan kesesuaian harga dengan manfaat.

Riset yang kedua [6], dilakukan oleh perusahaan mobil di Shanghai, China. Berpendapat bahwa harga yang terjangkau membuat pembeli maupun penjual melakukan transaksi dengan tenang. Dengan populasi hampir 100.000 dari data pembelian dealer mobil di Negara China. Perusahaan mobil ini, menggunakan linear regresi dan *random forest*, dengan memisahkan menjadi 3 model, pertama, *a model of certain car make*, yaitu semua sampel yang digunakan untuk membangun mobil dengan merek dan tahun pembuatan yang sama, kedua, *a model for a certain car series*, yaitu sampel untuk seri mobil tertentu, ketiga, *a universal model*, yaitu sampel untuk seluruh populasi mobil.

Riset yang ketiga [12], penelitian ini mempertimbangkan permintaan mobil pribadi bekas yang sudah meningkat dan menjadi sebuah peluang bisnis yang baik bagi pembeli maupun penjual. Jurnal ini menggunakan *multiple linear regression*, *random forest regression* dan *gradient boosted regression* untuk membangun sebuah harga model dari mobil bekas. Penelitian ini memanfaatkan regresi linear dengan melakukan perbandingan terhadap nilai koefisien korelasi (R) dari setiap atribut yang ada.

III. METHODOLOGY

Tiga riset pada bab sebelumnya dapat dijadikan landasan untuk memulai riset kami sendiri mengenai hubungan antara produksi mobil dengan harganya.

A. Dataset

Untuk menganalisis harga produk terhadap harga pasar, kami telah mengambil data Nigerian Car Price, yang digunakan sebagai acuan dalam penelitian. Data set ini diambil dari salah satu perusahaan pasar online yang menghubungkan penjual dan pembeli mobil bekas.

Dalam dataset ini, response variabelnya adalah harga mobil (Price) dan variable penjelasnya adalah tahun produksi mobil (Year). Untuk membandingkan efek regresi linear lebih baik, kami menggunakan 3 model regresi yang berbeda, yaitu Linear Regression Model, Loess Method, dan Robust Fitting of Linear Models Method.

B. Data Pre-Processing

1) Data Cleaning

Data Cleaning atau disebut juga pembersihan data merupakan bagian dari data Pre-Processing yang berguna untuk memastikan kebenaran, konsistensi, dan kegunaan suatu data sehingga data dapat di proses. Proses dalam data cleaning meliputi menghapus data atau mengisi data yang kosong, menghaluskan data, menghilangkan inkonsistensi sebuah data, mendeteksi adanya error atau nilai corrupt pada sebuah data dan memperbaiki kesalahan data tersebut, serta melakukan filter outlier yang tidak digunakan. Proses data cleaning bertujuan agar data dapat lebih mudah diolah, meminimalisir tingkat error dalam melakukan analisa, serta meningkatkan efisiensi dalam melakukan analisa data. Proses data cleaning dapat digunakan nantinya dalam melakukan regresi.

2) Data Reduction

Data reduction sebagai bagian dari proses data Pre-Processing, yaitu sebagai process pengolahan data yang biasanya dikumpulkan peneliti dalam melakukan penelitian di lapangan sebelum digunakan sebagai laporan akhir. Oleh karena itu, data reduction dapat diartikan sebagai suatu proses pemilihan data, penyederhanaan data, dan transformasi dari data yang masih kasar yang dihasilkan dari hasil pengamatan di lapangan untuk mencapai tujuan dari penelitian. Proses data reduction bertujuan untuk menghasilkan laporan akhir yang sudah tersusun dengan baik dan berkualitas.

C. Exploratory Data Analysis

1) Uji Asumsi

Melakukan uji asumsi dengan memanfaatkan Uji Asumsi Klasik. Pengujian Asumsi Klasik merupakan pengujian yang modelnya menerapkan adanya asumsi parameter populasi dari suatu sampel data penelitian yang harus memenuhi persyaratan dalam analisis regresi linear berganda berbasis OLS. Hal ini bertujuan untuk memastikan dataset kita merupakan model terbaik dalam hal konsistensi dan data tidak bias.

Untuk memenuhi tujuann dari Uji Asumsi Klasik, menggunakan 4 jenis pengujian, yaitu Pertama, Uji Normalitas Residul dengan menggunakan Shapiro-Wilk Normality Test, kedua, Uji Autokorelasi dengan

menggunakan Durbin-Watson Test, ketiga, Uji Homogenitas dengan menggunakan Breusch-Pagan Test, dan terakhir, keempat, Uji Multikolinearitas dengan acuan nilai VIF (*Variance Inflation Factor*).

2) Grubb's Outlier Test

Grubb's Test atau yang dikenal dengan Maximum Normalized Residual Test atau Extreme Studentized Deviate Test, merupakan pengujian yang dilakukan untuk mencari adanya outlier dalam sebuah univariate data dari populasi normal distribusi. Untuk melakukan Grubb's Outlier Test perlu didasari dengan asumsi normalitas. Oleh karena hal tersebut, sebelum melakukan pengujian outlier data harus dipastikan berdistribusi normal ketika dilakukan pendekatan

3) Regression Models

a) Linear Regression

$$y = \beta_0 + \beta_1 x$$

Jika kita memasukkan variable yang ada pada dataset, maka:

$$Price = -1.990e^{(9)} + 9.935e^{(5)} * Year$$

Linear model bergantung pada beberapa asumsi. Asumsi ini akan krusial dalam menguji validitas dan ketergunaan dari suatu linear model. Jika asumsi – asumsi ini akan menjadi fondasi kita dalam menggunakan linear model. Pertama, Linearitas mengasumsikan bahwa variable yang digunakan dalam linear regression harus membentuk sebuah garis. Normalitas mengasumsikan bahwa data yang digunakan terdistribusi secara normal. Independence mengasumsikan error yang terjadi, terjadi tersendiri. Homoscedasticity mengasumsikan error terdistribusi kepada semua variable prediksi. Jika beberapa asumsi ini tidak terpenuhi, interpretasi dari data harus lebih waspada [7].

b) Loess Method for Linear Regression

Metode LOESS menghitung prediksi yang dihaluskan dengan menyesuaikan model regresi lokal ke data. Berikut merupakan langkah dasar perhitungan metode LOESS:

1. Untuk setiap titik data, fungsi pembobotan diterapkan ke titik data lain dalam kumpulan data. Fungsi pembobotan memberikan bobot ke setiap titik data berdasarkan jaraknya ke tempat tujuan.
2. Model regresi linier tertimbang kemudian disesuaikan dengan titik data terdekat. Bobot yang diberikan ke setiap titik data dalam fungsi pembobotan menentukan tingkat kepentingan relatif setiap titik data dalam model regresi.
3. Model yang dipasang digunakan untuk memprediksi nilai variabel respons pada titik perhatian.
4. Proses ini kemudian diulang untuk setiap titik dalam dataset.
5. Parameter pemulusan, biasanya dilambangkan dengan "span" yang mengatur pengaruh setiap titik, span yang kecil akan menghasilkan model yang lebih fleksibel dan span yang lebih besar akan menghasilkan model yang lebih halus.

Hasil akhirnya adalah kurva halus yang sesuai dengan data dengan baik dan dapat digunakan untuk prediksi atau untuk mengidentifikasi tren atau pola dalam data.

c) Robust Method for Linear Regression

Perhitungan metode Robust melibatkan beberapa langkah, tergantung pada metode spesifik yang digunakan. Namun secara umum, langkah dasarnya adalah:

1. mengidentifikasi titik data yang dianggap outlier dan menghapusnya dari kumpulan data. Hal ini dilakukan untuk memastikan bahwa hasil akhir tidak terpengaruh oleh nilai ekstrim yang dapat memiringkan hasil.
2. Lokasi dan skala data diperkirakan, biasanya dengan menghitung median dan rentang interkuartil (IQR) untuk data tersebut.

Robust estimator adalah estimasi statistik yang kurang sensitif terhadap keberadaan outlier dalam data. Beberapa Robust estimator yang umum termasuk median, trimmed mean, dan M-estimator.

D. Measurements

1. R squared
Nilai R-squared harus berada di antara 0 dan 1. Nilai 0 menunjukkan bahwa variabel respon X sama sekali tidak dapat dijelaskan oleh variabel prediktor Y.

Nilai 1 menunjukkan bahwa variabel respon X dapat dijelaskan dengan sempurna tanpa kesalahan oleh variabel prediktor Y.
2. Adjusted R squared
Ini adalah versi modifikasi dari R-squared yang telah disesuaikan dengan jumlah prediktor dalam model. Versi modifikasi dari R-squared selalu lebih rendah dari R-kuadrat. R-kuadrat yang disesuaikan dapat berguna untuk membandingkan kecocokan model regresi yang berbeda satu sama lain.
3. P- Value
P-value menunjukkan apakah hipotesis dapat diterima atau ditolak. Hipotesis dalam hal ini, adalah nilai prediktor tidak bermakna untuk model linear Regresinya. Cara standar untuk menguji apakah prediktor tidak bermakna adalah dengan melihat apakah nilai p lebih kecil dari 0,05 (alpha).
4. Standard Error
Standard Error regresi memberikan ukuran absolut dari jarak titik data yang jatuh terhadap garis regresi. Semakin rendah nilai Standard Error menandakan bahwa jarak antara titik data dan nilai pas semakin kecil.

Kesalahan standar regresi berguna karena dapat digunakan untuk menilai ketepatan prediksi. Kira-kira 95% dari pengamatan harus berada dalam +/- dua kali kesalahan standar regresi.

5. Mean Squared Error

Mean Square Error (MSE) memberikan nilai seberapa dekat nilai prediksi dengan nilai sebenarnya. MSE ini didapatkan dari selisih rata-rata kuadrat antara nilai prediksi dengan nilai sebenarnya.

$$\text{Mean Square Error} = \frac{1}{n} \times \sum (x_1 - x_2)^2$$

Dimana x_1 = Nilai prediksi, x_2 = Nilai sebenarnya, n = jumlah titik data

Apabila nilai MSE semakin kecil maka semakin dekat nilai prediksi dengan nilai sebenarnya. Oleh karena itu, MSE berguna untuk mengetahui kedekatan antara kedua variabel tersebut.

IV. RESULTS

A. Data Pre-Processing

1) Data Cleaning

a) Menghapus kolom "Build" dari dataset

Hal ini dilakukan karena pada dataset bagian "Build" memiliki nilai NA sebesar 72.48% dari jumlah data kami. Oleh karena itu, kami tidak menggunakan kolom tersebut karena data yang memiliki nilai NA begitu banyak.

b) Menghapus Data yang Memiliki Nilai NA

Setelah kami menghapus kolom bagian "Build", kami menghitung berapa banyak data yang memiliki NA. Dari kode kita yang jalankan terdapat 2047 data yang memiliki nilai NA. Oleh karena itu, kami menghapus 2047 data tersebut dari data frame yang kami gunakan untuk analisis.

c) Mengubah Nama Kolom

Ada 2 kolom yang kita ubah namanya seperti "Year of manufacture" dan "Engine size". Kami mengubah nama kolom tersebut menjadi "Year" dan "Engine" karena nama kolom asli mengandung spasi yang dapat menyulitkan kami dalam menjalankan kode.

d) Hasil Data yang Telah Dibersihkan

Setelah data kami bersihkan, kami memiliki 1101 data dan 8 kolom yang dapat kami gunakan untuk menganalisis dataset tersebut.

B. Exploratory Data Analysis

1) Uji Asumsi

a. Uji Normalitas

Pengujian Normalitas dari data set yang ada dapat menggunakan Shapiro-Wilk normality test. Metode ini merupakan metode pengujian normalitas suatu data yang efektif. Metode ini bertujuan untuk membuktikan apakah nilai residu berdistribusi normal atau tidak. Untuk mengujinya, kita memanfaatkan nilai alpha (0.05 / 5%) dan nilai p-value, dengan ketentuan:

H_0 = Data Terdistribusi Normal	
P-value < 0.05	Menolak Hipotesa 0
P-value > 0.05	Menerima Hipotesa 0

Menggunakan Shapiro-Wilk normality test, harus memenuhi 3 syarat:

1. Data berskala interval / ratio (kuantitatif)
2. Data tunggal / belum dikelompokkan ke dalam tabel Distribusi frekuensi
3. Data dari sampel random

Hasil dari tes Shapiro-Wilk menunjukkan bahwa $P - Value = 2.2e - 16$.

Dengan demikian, H_0 ditolak dan data tidak terdistribusi normal.

b. Uji Autokorelasi

Pengujian Autokorelasi nilai residual dari sebuah analisis regresi dapat menggunakan Durbin-Watson test. Metode ini dapat mendeteksi terjadinya autokorelasi dari suatu data dengan membuktikan perhitungan p-value dan nilai Durbin-Watson. Untuk pengujian Durbin-Watson test dapat melihat DU dan DL (dari tabel Durbin-Watson) untuk menentukan autokorelasi positif ataupun autokorelasi negatif, dengan ketentuan:

$$H_0 = \text{Data terjadi korelasi}$$

DW < DL	Autokorelasi positif
DW > DU	Autokorelasi Negatif
DL < DW < DU	Tidak dapat disimpulkan

Dari hasil uji autokorelasi menggunakan Durbin-Watson menghasilkan $P - Value = 0.547$. Terlihat bahwa p-value > alpha (0.05) sehingga dalam data tersebut tidak terjadi autokorelasi.

c. Uji Homogenitas

Pengujian homogenitas dari data set dapat menggunakan Breusch-Pagan test. Metode ini membuktikan homogenitas suatu data residual dengan melihat p-value. Untuk setiap data yang homogen, berarti memiliki nilai varians residual yang sama, sehingga varians residual bersifat konstan. Pembuktian menggunakan metode Breusch-Pagan test ini memiliki ketentuan:

$$H_0 = \text{Data mempunyai Homogenitas}$$

P-value < 0.05	Menolak Hipotesa 0
P-value > 0.05	Menerima Hipotesa 0

Dari hasil uji homogenitas menggunakan Breusch-Pagan test terlihat bahwa p-value < alpha (0.05) sehingga terjadi homogenitas dalam data tersebut. Karena terjadi homogenitas dalam data tersebut, nilai variansi data-data tersebut sama.

d. Uji Multikolinearitas

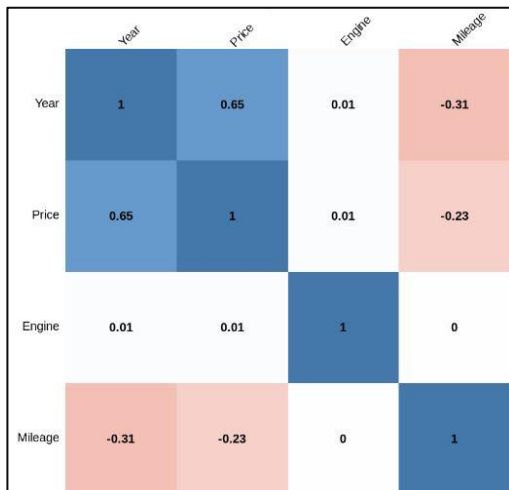
Pengujian multikolinearitas dari suatu data dapat menggunakan nilai VIF. VIF (Variance Inflation Factor) merupakan ukuran jumlah multikolinearitas dalam suatu set variabel (variabel independent) dari regresi linear. Untuk membuktikan terjadinya multikolinearitas dengan metode ini, memiliki ketentuan:

VIF < 10	Tidak terdapat masalah multikolinearitas
VIF > 10	Terdapat masalah multikolinearitas

VIF yang dihasilkan untuk variable *Year* = 1.1102.

Dari hasil uji multikolinearitas diatas, terlihat bahwa setiap nilai VIF < 10, sehingga dalam data tersebut tidak ada masalah multikolinearitas

2) *Exploratory Data Analysis*

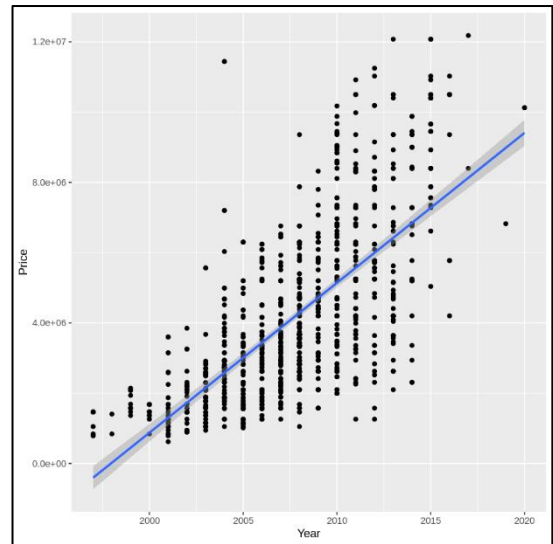


Gambar 4.2 Correlation Plot

Dari Gambar 4.1 variabel *Year* dan *Price* mempunyai korelasi koefisien yang bernilai 0.65. Angka ini menandakan bahwa harga mobil akan bertambah seiring dengan bertambah tahun produksinya.

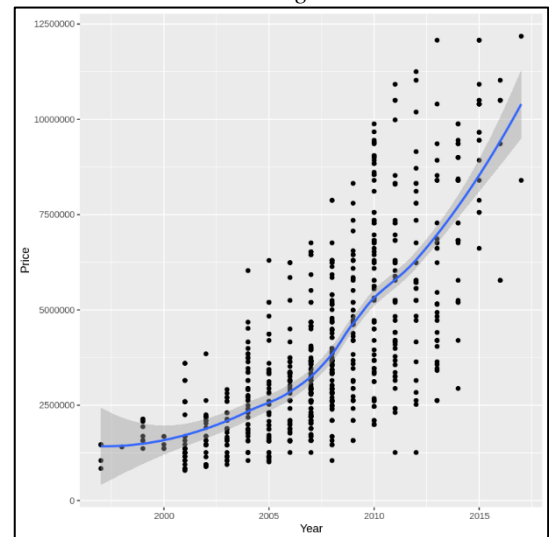
C. Regressions

1) *Linear Regression*



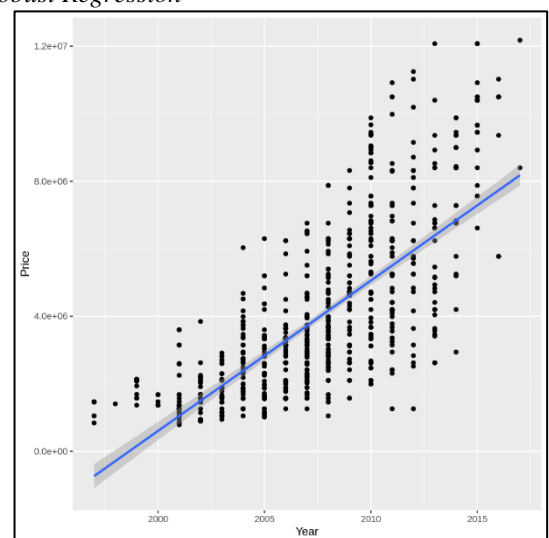
Gambar 4.3 Linear Regression

2) *Loess Method For Linear Regression*



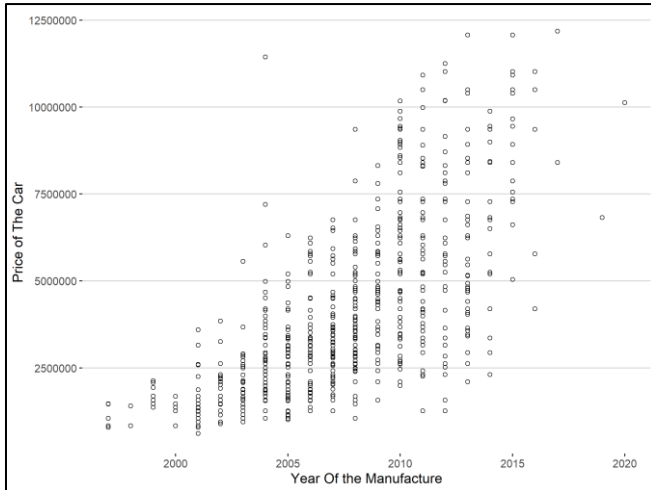
Gambar 4.4 Loess Method

3) *Robust Regression*



Gambar 4.4 Robust Regression

V. ANALYSIS AND DISCUSSION



Gambar 5.1 Scatterplot Harga Mobil dengan Tahun Pembuatannya

TABLE I. SUMMARY LINEAR MODEL

Residuals				
Min	1Q	Median	3Q	Max
-4852811	-1231889	-144922	977925	5497848

Coefficients				
	Estimate	Std. Error	t value	Pr(> t)
Intercept	-928141676	32589676	-28.48	<2e-16***
Year	464341	16232	28.61	<2e-16***

SIGNIF. CODES: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

RESIDUAL STANDARD ERROR: 1710000 ON 768 DEGREES OF FREEDOM

MULTIPLE R-SQUARED: 0.5159

ADJUSTED R-SQUARED: 0.5152

F-STATISTIC: 818.3 ON 1 AND 768 DF

P-VALUE: < 2.2E-16

Hasil diatas didapatkan dengan menjalankan kode `summary()` yang dipadukan dengan metode `lm()`. Hasil tersebut menampilkan linear regresion dari semua variable.

TABLE II. SUMMARY LOESS METHOD

Number of Observation	770
Equivalent Number of Parameters	5.15
Residual Standard Error	1635000
Trace of Smoother Matrix	5.63 (exact)

Hasil diatas didapatkan dengan menjalankan kode `loess()` terhadap variabel Price dengan Year. Setelah itu dimasukkan ke fungsi `summary()` untuk mendapatkan semua hasil yang dikeluarkan oleh metode LOESS.

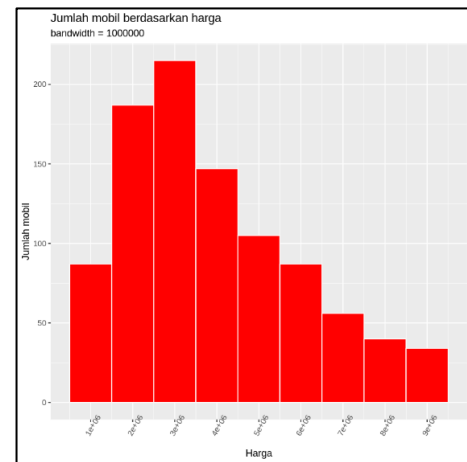
TABLE III. SUMMARY ROBUST METHOD

Residuals				
Min	1Q	Median	3Q	Max
-4690346	-1146275	-75247	1066028	5679016

Coefficients			
	Value	Std. Error	t Value
(Intercept)	-8.906721e+08	3.202732e+07	-2.780980e+01
Year	4.456374e+05	1.595209e+04	2.793600e+01

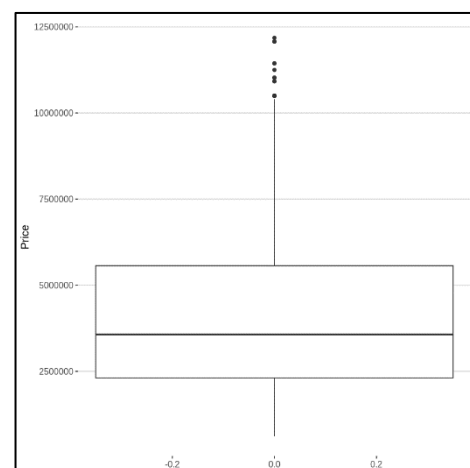
Residual standard error: 1666000 on 768 degrees of freedom

Hasil diatas didapatkan dengan menjalankan kode `rm()` terhadap variabel Price dengan Year. Setelah itu dimasukkan ke fungsi `summary()` untuk mendapatkan semua hasil yang dikeluarkan oleh metode Robust.



Gambar 5.2 Distribution Price

Berdasarkan dataset yang sudah dibersihkan, harga mobil mempunyai kemiringan ke kanan. Akibatnya, nilai mobil terkelompok di daerah yang kurang dari rata-rata-ratanya. Ini menandakan ada beberapa yang data yang berada di luar distribusi normal.



Gambar 5.3 Boxplot

Dari gambar 5.3, dapat dilihat outlier berada di luar interval rata-rata. Oleh karena itu, diperlukan regresi yang fleksibel dalam menangani outlier. Fleksibilitas ini dapat dilihat dari Residual Standard Error. Residual Standard Error yang telah

didapatkan dari Table I, II, dan III. Residual Standard Error menandakan keefektifitas suatu model dalam memprediksi suatu nilai jika ada outliernya.

TABLE IV. RESIDUAL STANDARD ERROR

Residual Standard Error	
Linear Model	1710000
Loess Method	1635000
Robust Method	1666000

Berdasarkan Residual Standard Error, Loess method menghasilkan Residual Standard Error yang paling kecil. Dengan demikian data yang diprediksi akan mempunyai perbedaan yang paling kecil terhadap data aslinya.

TABLE V. T-VALUE OF LINEAR MODEL AND ROBUST METHOD

T-value to Year	
Linear Model	28.610
Robust Method	27.936

T-value menandakan seberapa jauh hasil percobaan dengan kondisi umumnya. Semakin besar nilai T-nya, semakin besar bukti yang melawan hipotesis nol. Dimana H_0 merupakan kemampuan Year dapat memprediksi Price. Dari Linear Model dengan Robust Method, nilai t-value dari Robust Method merupakan nilai yang lebih kecil dibandingkan Linear Model.

VI. CONCLUSION

Transportasi merupakan hal yang esensial dalam kehidupan modern. Salah satu transportasi yang banyak kita jumpai yaitu mobil. Keputusan untuk membeli suatu mobil merupakan perihal yang perlu pertimbangan yang matang. Dengan demikian, model-model regresi harus akurat dalam memprediksi harga mobil yang sesuai dengan tahun produksinya.

Setiap model regresi memiliki keuntungan dan kerugiannya masing-masing yang secara singkat dapat dipaparkan sebagai berikut.

Beberapa keuntungan dari penggunaan Linear Model yaitu:

1. Menghasilkan summary yang lebih lengkap
2. Efisiensi waktu dalam melakukan prediksi

Beberapa kerugian dari penggunaan Linear Model yaitu:

1. Kurang efektif terhadap data yang tidak terdistribusi normal
2. Keterbatasan dalam mengatasi data yang memiliki outlier

Beberapa keuntungan dari penggunaan metode LOESS yaitu:

1. Efektivitas penanganan data tidak teratur atau cenderung berisiko.
2. Kemampuan untuk menghitung prediksi dengan tingkat ketelitian yang tinggi.
3. Kemampuan untuk menangani data dengan jumlah variabel yang besar.

4. Kemampuan untuk menangani data dengan jumlah sampel yang kecil.

Sedangkan berikut beberapa kerugian dari penggunaan metode LOESS yaitu:

1. Metode ini tidak dapat digunakan untuk menangani data dengan jumlah sampel yang besar karena akan memakan banyak waktu dan sumber daya.
2. Metode ini mungkin tidak stabil dalam menangani data dengan tingkat variasi yang tinggi.
3. Metode ini mungkin tidak cocok untuk digunakan dalam situasi dimana interpretasi model diperlukan.
4. Dalam beberapa kasus, metode ini mungkin tidak dapat digunakan untuk menangani data dengan jumlah variabel yang sangat besar.

Selanjut mengenai keuntungan dari penggunaan metode Robust yaitu:

1. Efektivitas penanganan data yang memiliki outlier (data yang sangat tidak sesuai dengan kebanyakan data lainnya).
2. Mendapatkan hasil yang stabil dan tidak terpengaruh oleh outlier.
3. Dapat menangani data dengan varians yang tidak konstan.
4. Kemampuan untuk menangani data dengan jumlah sampel yang kecil.

Sedangkan berikut beberapa kerugian dari penggunaan metode Robust yaitu:

1. Metode ini mungkin akan menghasilkan hasil yang kurang akurat dibandingkan dengan metode tradisional jika data tidak memiliki outlier.
2. Metode ini mungkin memerlukan lebih banyak waktu dan sumber daya untuk dilakukan.
3. Metode ini mungkin tidak dapat digunakan untuk menangani data dengan jumlah variabel yang sangat besar.
4. Metode ini mungkin tidak dapat digunakan untuk menangani data dengan tingkat kompleksitas yang tinggi. Pemilihan metode robust yang tepat dapat menjadi sulit dan memerlukan pengetahuan yang cukup dalam bidang statistika.

Dengan mempertimbangkan keuntungan dan kerugian dari setiap model yang ada, model yang paling cocok untuk memprediksi harga mobil berdasarkan tahunnya merupakan regresi LOESS method. Pertimbangan pertama, kemampuan LOESS method untuk memprediksi data yang memiliki banyak outliernya. Hal ini didasarkan dari Residual Standard Error LOESS method yang relatif kecil. Pertimbangan kedua, data yang digunakan memiliki jumlah variabel yang relatif banyak dan jumlah sample yang tidak begitu banyak.

Dalam penelitian ini, peneliti menemukan beberapa keterbatasan terutama pada pemilihan model regresi dan pengambilan variabel. Kesimpulan yang dapat ditarik yakni untuk optimalisasi penelitian selanjutnya dapat dilakukan dengan analisis lebih lanjut terhadap harga mobil dengan menyertakan beberapa variabel independen lain, seperti kekuatan mesin, kondisi mobil, dan jumlah jarak tempuh serta model regresi lainnya.

REFERENCES

- [1] NIST. (2012). *e-Handbook of Statistical Methods*. Retrived from <http://www.itl.nist.gov/div898/handbook/>
- [2] Iqbal, M. (2015). Regresi Data Panel (2): Tahap Analisis. Retrived from <https://dosen.perbanas.id/regresi-data-panel-2-tahap-analisis>.
- [3] Gerung, C. J., Sepang, J., & Loindong, S. (2017). Pengaruh Kualitas Produk, Harga Dan Promosi Terhadap Keputusan Pembelian Mobil Nissan X-Trail Pada Pt. Wahana Wirawan Manado. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*, 5(2).
- [4] Chen, C., Hao, L., & Xu, C. (2017, May). Comparative analysis of used car price evaluation models. In *AIP Conference Proceedings* (Vol. 1839, No. 1, p. 020165). AIP Publishing LLC. (Vol. 1839, No. 1, p. 020165). AIP Publishing LLC. Retrived from <https://aip.scitation.org/doi/abs/10.1063/1.4982530>
- Dodge, Y. (1997). LAD Regression for Detecting Outliers in Response and Explanatory Variables. Retrived from <https://www.sciencedirect.com/science/article/pii/S0047259X97916663>
- [5] Ozgur, C., Parveen, S., et al. (2016). Multiple Linear Regression Applications in Automobile Pricing. Retrived from https://www.researchgate.net/profile/Ceyhun-Ozgur-2/publication/306058279_C_Ozgur_ZHughes_G_Rogers_S_Parveen_Multiple_Linear_Regression_Applications_in_Automobile_Pricing_submitted_to_International_Journal_of_Mathematics_and_Statistics_Invention_vol_4_no_6_August_2016_pp1-1/links/57ace21d08ae7a6420c32b57/C-Ozgur-ZHughes-G-Rogers-S-Parveen-Multiple-Linear-Regression-Applications-in-Automobile-Pricing-submitted-to-International-Journal-of-Mathematics-and-Statistics-Invention-vol-4-no-6-August-2016.pdf
- [6] Hickey GL, Kontopantelis E, Takkenberg JJM, Beyersdorf F. (2019). Statistical primer: checking model assumptions with regression diagnostics. *Interact CardioVasc Thorac Surg*;28:1–8.
- [7] Breslow, N. E. (1995). *Generalized Linear Models: Checking Assumptions and Strengthening Conclusions*. Retrived from <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=202246e2d6d09f0804435b3e66c0c716865336a3>
- [8] Aggarwal R., Ranganathan P.(2017). Common pitfalls in statistical analysis: Linear regression analysis. *Perspect Clin Res*. 2017 Apr-Jun;8(2):100-102. doi: 10.4103/2229-3485.203040. PMID: 28447022; PMCID: PMC5384397.
- [9] Kodu, S. (2013). Harga, kualitas produk dan kualitas pelayanan pengaruhnya terhadap keputusan pembelian mobil Toyota avanza. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, 1(3).
- [10] Wahyuni, Silvia T., (2019). PEMODELAN GENERALIZED ADDITIVE MODEL FOR LOCATION, SCALE, AND SHAPE (GAMLSS) DENGAN PEMULUSAN LOCALLY ESTIMATED SCATTERPLOT SMOOTHING (LOESS) PADA KASUS HIV/AIDS DI JAWA TIMUR. Retrived from <http://repository.unimus.ac.id/3944/8/JURNAL.pd>
- [11] Monburinon, N., Chertchom, P., Kaewkiriya, T., Rungpheung, S., Buaya, S., & Boonpou, P. (2018, May). Prediction of prices for used car by using regression models. In *2018 5th International Conference on Business and Industrial Research (ICBIR)* (pp. 115-119). IEEE.