

Model Prediksi Harga Mobil Bekas Menggunakan Regresi Linear Berganda Berdasarkan Fitur Kendaraan

Calvin Whenjaya¹, Hadi Susanto^{2*}, Herman³
^{1,2,3}Teknik Perangkat Lunak, Universitas Universal
*Corresponding author E-mail: hadi.s256314@uvers.ac.id

Article Info

Article history:

Received 02-12-2024
Revised 10-12-2024
Accepted 28-12-2024

Keyword:

Mobil Bekas, Jenis Sparepart,
Brand Mobil, Regresi Linear

ABSTRACT

Used cars are one of the popular segments in the car sales market in Indonesia today. This study aims to understand the used car market in Indonesia in depth. Through data analysis and surveys, this research collects data containing information on various used car listings, their attributes and features, including brand, manufacturing year, price, installment amount, mileage, transmission type, location, license plate type, and various features such as rear cameras, sunroof, automatic side mirrors, and others. The method used in this study is multiple linear regression, which aims to model the relationship between the price of used cars and several predictor variables. By using this method, the study can analyze how each factor, such as brand, manufacturing year, and other features, influences the price of used cars offered in the Indonesian market. This multiple linear regression model helps to identify and predict the price of used cars based on the combination of various factors, providing a more accurate picture of the dynamics of the used car market. In general, this study provides a comprehensive overview of predicting the price of used cars based on their features in Indonesia.

Copyright © 2024 Journal of Digital Ecosystem for Natural Sustainability.
All rights reserved.

I. PENDAHULUAN

Transportasi memegang peranan penting dalam kehidupan masyarakat modern, khususnya dalam mendukung mobilitas individu maupun kegiatan ekonomi. Salah satu aspek yang menjadi perhatian adalah meningkatnya kebutuhan kendaraan pribadi di tengah terbatasnya akses terhadap transportasi umum yang nyaman dan efisien. Fenomena ini memicu tingginya permintaan pasar untuk kendaraan bekas sebagai alternatif ekonomis dibandingkan dengan kendaraan baru. Namun, dinamika harga kendaraan bekas sering kali tidak transparan dan sulit diprediksi, sehingga menjadi tantangan tersendiri bagi konsumen dan pelaku pasar [1].

Salah satu permasalahan yang sering dihadapi masyarakat dalam memenuhi kebutuhan transportasi adalah pemilihan mobil bekas. Dengan beragamnya pilihan merek, model, tahun produksi, dan harga, konsumen seringkali kesulitan dalam mengambil keputusan yang tepat. Ketidakpastian mengenai kondisi kendaraan, potensi masalah teknis, dan harga jual yang fluktuatif menjadi kendala utama. Hal ini

membutuhkan analisis data yang mendalam untuk membantu konsumen dalam membuat keputusan yang lebih baik [2].

Beberapa penelitian sebelumnya telah berhasil mengembangkan metode prediksi untuk menyelesaikan permasalahan serupa. Misalnya, penelitian oleh Smith pada tahun 2020 menunjukkan bahwa regresi linear dapat digunakan untuk memprediksi harga pasar kendaraan bermotor berdasarkan atribut seperti usia kendaraan dan kondisi fisik. Penelitian lain oleh Zhang pada tahun 2021 menerapkan regresi linear berganda untuk memperkirakan harga pasar properti dengan mempertimbangkan berbagai variabel seperti lokasi dan ukuran. Kedua penelitian ini menunjukkan bahwa metode prediksi berbasis regresi dapat memberikan hasil yang akurat untuk masalah prediktif dengan variabel kuantitatif [3].

Regresi linear adalah salah satu metode analisis statistik yang digunakan untuk menentukan hubungan antara satu atau lebih variabel independen dengan variabel dependen. Metode ini memiliki kelebihan dalam kesederhanaan implementasi, kemampuan interpretasi, dan efisiensi dalam pemrosesan

data. Menurut Gujarati dan Porter pada tahun 2017, regresi linear cocok untuk menganalisis data dengan pola hubungan linear yang jelas, sehingga relevan digunakan dalam prediksi harga mobil bekas. Selain itu, regresi linear dapat dikembangkan menjadi regresi berganda untuk menangani variabel yang lebih kompleks, menjadikannya alat yang fleksibel dalam berbagai kasus prediksi [4].

Penelitian ini bertujuan untuk prediksi selisih harga mobil bekas di Indonesia yang berasal dari *website carsome.id* *website* ini adalah penyedia informasi untuk mobil bekas di Indonesia. Dengan menggunakan regresi linear, model ini akan memanfaatkan data historis harga mobil bekas dan atribut terkait untuk mengidentifikasi pola yang mempengaruhi harga. Proses pengerjaan melibatkan pengumpulan data, eksplorasi data, pengembangan model regresi linear, serta evaluasi performa model. Dengan hasil yang diperoleh, diharapkan model ini dapat memberikan rekomendasi harga yang lebih akurat bagi konsumen dan pelaku pasar di Indonesia.

II. METODE

Penelitian ini menggunakan metode regresi linear, Dengan tujuan untuk memprediksikan harga mobil bekas di Indonesia dengan menyertakan beberapa spesifikasi dan komponen dari mobil tersebut.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan hasil data yang didapatkan dari <https://www.kaggle.com/datasets/indraputra21/used-car-listings-in-indonesia>. Data tersebut *dipublish* oleh Indra Putra. Datasetnya tersusun atas 8 variabel utama dan 12.500 baris data. Data ini dipilih karena kelengkapan atribut yang relevan untuk analisis prediksi harga mobil bekas.

2.2 Data Preprocessing

Proses preprocessing dilakukan untuk memastikan data siap digunakan dalam model regresi linear berganda. Langkah-langkah yang dilakukan meliputi:

Perubahan ke Faktor: Bagaimana data berubah atau bergeser seiring dengan perubahan faktor tertentu. Faktor ini bisa berupa variabel independen yang peneliti manipulasi atau ukur untuk melihat dampaknya terhadap data (variabel dependen).

Pemilihan Variabel: Variabel yang digunakan untuk prediksi ditampilkan dalam tabel berikut:

Tabel 1. Deskripsi Variabel X dan Y

| Variabel | Deskripsi | Tipe Data | Satuan |
|--------------|-----------------------|---------------|--------|
| <i>Brand</i> | Merek atau manufaktur | <i>String</i> | - |

| | dari mobil. | | |
|----------------------------------|--|----------------|---|
| <i>Transmission</i> | Tipe dari transmisi mobil seperti manual atau otomatis. | <i>String</i> | - |
| <i>Rear Camera</i> | Indikasi apakah ada kamera depan atau tidak. | <i>Boolean</i> | - |
| <i>Sun Roof</i> | Indikasi apakah ada <i>sunroof</i> atau tidak | <i>Boolean</i> | - |
| <i>Auto Retract Mirror</i> | Indikasi apakah ada fitur spion lipat atau tidak. | Boolean | - |
| <i>Electric Parking Brake</i> | Indikasi apakah ada rem parkir elektrik, | Boolean | - |
| <i>Map Navigator</i> | Indikasi apakah ada navigasi peta atau tidak. | Boolean | - |
| <i>Vehicle Stability Control</i> | Indikasi apakah dilengkapi dengan kontrol keseimbangan mobil. | Boolean | - |
| <i>Keyless Push Start</i> | Indikasi apakah ada fitur tombol tekan untuk menghidupkan mesin atau tidak | Boolean | - |
| <i>Sports Mode</i> | Indikasi apakah ada mode <i>sport</i> atau tidak. | Boolean | - |
| <i>360 Camera View</i> | Indikasi apakah ada kamera 360° atau tidak. | Boolean | - |
| <i>Power Sliding Door</i> | Indikasi apakah pintu geser atau tidak. | Boolean | - |
| <i>Auto Cruise Control</i> | Indikasi apakah fitur | Boolean | - |

| | | | |
|--------------|--|----------------|--------|
| | <i>cruise control</i> ada atau tidak. | | |
| <i>Price</i> | Harga dari mobil tersebut | <i>Integer</i> | Rupiah |

2.3 Pemrograman R

R merupakan sebuah bahasa pemrograman yang memungkinkan pengguna untuk memprogramkan algoritma dengan alat yang dikembangkan melalui R oleh pengguna lainnya oleh Zuur pada tahun 2009. R merupakan bahasa pemrograman tingkat tinggi untuk analisis data dan grafik [5].

2.4 Metode Regresi Linear

Regresi linear adalah metode analisis yang digunakan untuk memprediksi variabel dependen (harga mobil bekas) berdasarkan satu atau lebih variabel independen (fitur kendaraan). Model regresi linear berganda dinyatakan dalam persamaan:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Gambar 1. Rumus regresi linear pada gambar

Di mana:

- Y : Variabel dependen (harga mobil bekas).
- $X_1...X_n$: Variabel independen (atribut kendaraan).
- β_0 : Koefisien *Intercept*.
- $\beta_1... \beta_n$: Koefisien regresi.
- ϵ : Error atau residual.

Metode ini diasumsikan memiliki hubungan linear antara variabel dependen dan independen serta variabel residual yang mengikuti distribusi normal. Estimasi parameter dilakukan dengan metode **ordinary least squares (OLS)** untuk meminimalkan jumlah kuadrat error [6].

2.5 Evaluasi Prediksi

Untuk mengevaluasi akurasi model regresi linear, digunakan dua metrik utama:

Koefisien Determinasi (R^2)

Koefisien determinasi (R^2) mengukur proporsi variabilitas variabel dependen (Y) yang dapat dijelaskan oleh kombinasi variabel independen (X). Nilai R^2 berada dalam rentang 0 hingga 1:

$R^2 = 1$: Model menjelaskan semua variabilitas

dalam data.

$R^2 = 0$: Model tidak menjelaskan variabilitas data sama sekali.

Rumus R^2 :

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

$\sum(y_i - \hat{y}_i)^2$: Residual Sum of Square (RSS), kesalahan prediksi

$\sum(y_i - \bar{y})^2$: Total Sum of Squares (TSS), total variabilitas data aktual.

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

RMSE memberikan penalti yang lebih besar pada error yang besar dibandingkan MAE. Kedua metrik ini digunakan untuk membandingkan performa model dengan tujuan meminimalkan nilai MAE dan RMSE sehingga model memiliki kemampuan prediksi yang lebih baik [1].

III. HASIL DAN PEMBAHASAN

1. Import Dataset

Dalam pengembangan penelitian ini kami mengimport data yang tersedia pada *google sheet* ke dalam lingkungan pengembangan *google collaboration* dengan menggunakan fungsi *read_sheet()* yang tersedia dalam paket "googlesheets4".

```
install.packages("googlesheets4")
library(googlesheets4)

sheet_id <-
  "1hLjkYfL8mG4GAFtc1cuskf30Tfrx0vCJm2a3RNRmxo"
data <- read_sheet(sheet_id)
```

Data yang dikelola terdiri atas 1 Variabel dependen dan 15 variabel independen yang yang diubah kedalam bentuk factor dengan fungsi *as.factor()*.

```
data$brand <- as.factor(data$brand)
data$transmission <- as.factor(data$transmission)
data$rearCamera <- as.factor(data$rearCamera)
data$sunRoof <- as.factor(data$sunRoof)
data$autoRetractMirror <- as.factor(data$autoRetractMirror)
data$electricParkingBrake <-
```

```
as.factor(data$electricParkingBrake)
data$mapNavigator <- as.factor(data$mapNavigator)
data$vehicleStabilityControl <-
as.factor(data$vehicleStabilityControl)
data$keylessPushStart <- as.factor(data$keylessPushStart)
data$sportsMode <- as.factor(data$sportsMode)
data$`360CameraView` <- as.factor(data$`360CameraView`)
data$powerSlidingDoor <- as.factor(data$powerSlidingDoor)
data$autoCruiseControl <- as.factor(data$autoCruiseControl)
```

Dengan menentukan kategori referensi untuk variabel *brand* yaitu BMW dan *transmission* yaitu *Automatic* menggunakan fungsi *relevel()*.

```
data$brand <- relevel(data$brand, ref = "BMW")
data$transmission <- relevel(data$transmission, ref =
"Automatic")
```

2. Pemodelan Regresi Linear

Dalam Pemodelan ini, peneliti menggunakan teknik Regresi linear untuk menghubungkan antara variabel dependen (*priceRp*) dengan 15 variabel independen lainnya, teknik ini dikembangkan dengan menggunakan fungsi *lm()* di pemrograman R.

```
fit <- lm(data$priceRp ~ data$brand + data$year +
data$mileage + data$transmission + data$rearCamera +
data$sunRoof + data$autoRetractMirror +
data$electricParkingBrake + data$mapNavigator +
data$vehicleStabilityControl + data$keylessPushStart +
data$sportsMode + data$`360CameraView` +
data$powerSlidingDoor + data$autoCruiseControl)
```

Dengan model ini akan menghasilkan nilai-nilai koefisien regresi yang mewakili setiap variabel independen dengan hubungan terhadap variabel dependen.

| | |
|-----------------------------|---------------|
| (Intercept) | -9.883621e+09 |
| data\$brandChevrolet | -2.746740e+08 |
| data\$brandDaihatsu | -2.619897e+08 |
| data\$brandDatsun | -2.953067e+08 |
| data\$brandFord | -2.596163e+08 |
| data\$brandHonda | -2.455993e+08 |
| data\$brandHyundai | -1.832925e+08 |
| data\$brandMazda | -1.878564e+08 |
| data\$brandMitsubishi | -2.146467e+08 |
| data\$brandNissan | -2.527843e+08 |
| data\$brandSuzuki | -2.657369e+08 |
| data\$brandToyota | -2.238739e+08 |
| data\$brandWuling | -2.956805e+08 |
| data\$year | 5.102888e+06 |
| data\$mileage | -1.049589e+02 |
| data\$transmissionManual | -3.603846e+07 |
| data\$rearCamera1 | 1.552319e+07 |
| data\$sunRoof1 | 7.786110e+07 |
| data\$autoRetractMirror1 | 3.790163e+07 |
| data\$electricParkingBrake1 | 6.237640e+07 |
| data\$mapNavigator1 | 2.866441e+07 |

| | |
|--------------------------------|---------------|
| data\$vehicleStabilityControl1 | -3.684779e+06 |
| data\$keylessPushStart1 | 1.854030e+06 |
| data\$sportsMode1 | 3.216790e+06 |
| data\$`360CameraView`1 | 3.897552e+07 |
| data\$powerSlidingDoor1 | -2.627715e+07 |
| data\$autoCruiseControl1 | -3.662195e+06 |

3. Perancangan Regresi Linear Berganda

Perancangan regresi linear berganda melibatkan pengembangan model yang menghubungkan satu variabel dependen dengan lebih dari satu variabel independen. Dalam kasus ini, peneliti berfokus pada harga kendaraan (*priceRp*) sebagai variabel dependen yang dipengaruhi oleh sejumlah variabel independen seperti merek, tahun pembuatan, jarak tempuh, dan berbagai fitur lainnya.

```
predPriceRp <- coefficients[1] +
coefficients[brand] +
coefficients[14] * year +
coefficients[15] * mileage +
coefficients[16] * transmission +
coefficients[17] * rearCamera +
coefficients[18] * sunRoof +
coefficients[19] * autoRetractMirror +
coefficients[20] * electricParkingBrake +
coefficients[21] * mapNavigator +
coefficients[22] * vehicleStabilityControl +
coefficients[23] * keylessPushStart +
coefficients[24] * sportsMode +
coefficients[25] * camera360View +
coefficients[26] * powerSlidingDoor +
coefficients[27] * autoCruiseControl
```

Dengan model regresi linear berganda ini memungkinkan peneliti untuk memprediksi harga kendaraan berdasarkan berbagai faktor. Koefisien yang diperoleh dari model regresi linear memberikan gambaran yang jelas mengenai seberapa besar pengaruh setiap variabel terhadap harga kendaraan. Dengan menggunakan koefisien ini, peneliti dapat mengembangkan fungsi yang memungkinkan pengguna untuk memprediksi harga kendaraan hanya dengan memasukkan parameter-parameter.

4. Evaluasi Model

Evaluasi Model Regresi linear berganda sangat penting untuk mengukur keefektifan model dalam memprediksi hasil berdasarkan data aktual. Dua metrik utama yang sering digunakan untuk mengevaluasi kinerja model regresi adalah *R-squared* (R^2) dan *Root Mean Square Error* (RMSE).

```
alldata <- data.frame(data$brand, data$year, data$mileage,
data$transmission, data$rearCamera, data$sunRoof,
```

```

data$autoRetractMirror, data$electricParkingBrake,
data$mapNavigator, data$vehicleStabilityControl,
data$keylessPushStart, data$sportsMode,
data$`360CameraView`, data$powerSlidingDoor,
data$autoCruiseControl, data$priceRp)

predictions <- predict(fit, alldata)

# Menghitung RMSE

RMSE <- sqrt(mean((predictions - data$priceRp)^2))
cat("Root Mean Squared Error (RMSE):", RMSE, "\n")

# Menghitung R-squared

r_squared <- summary(fit)$r_squared
cat("R-squared:", r_squared, "\n")

```

Dengan metrik *R-square* (R^2) memberi tahu peneliti seberapa banyak variasi harga kendaraan yang dapat dijelaskan oleh model, sedangkan metrik RMSE memberi tahu peneliti seberapa akurat prediksi harga kendaraan dibandingkan dengan harga aktual.

Model regresi linear yang baik memiliki nilai R^2 yang tinggi (mendekati 1) dan RMSE yang rendah. Oleh karena itu, kedua metrik ini harus digunakan bersama-sama untuk mendapatkan gambaran yang lebih lengkap mengenai kinerja model.

```

Root Mean Squared Error (RMSE): 46798781
R-squared: 0.5095932

```

Nilai RMSE yang mencapai 46.798.781 disebabkan oleh tingginya varians data yang diolah, yang mencapai 66.882.598,57. Hal ini menunjukkan bahwa model kesulitan dalam memprediksi harga dengan akurat, karena adanya variasi yang besar dalam data. Sementara itu, nilai *R-squared* yang hanya mencapai 0,51 disebabkan oleh varians data yang cenderung fluktuatif, yang mengindikasikan bahwa model hanya mampu menjelaskan sepeneliti 51% variasi dalam harga kendaraan. Model ini mungkin tidak sepenuhnya mencakup faktor-faktor yang mempengaruhi harga secara optimal.

IV. KESIMPULAN

Penelitian ini bertujuan untuk memprediksi harga mobil bekas di Indonesia menggunakan metode regresi linear berganda. Model yang dikembangkan berhasil mengidentifikasi hubungan antara harga mobil dengan berbagai fitur seperti merek, tahun pembuatan, jarak tempuh, dan fitur tambahan. Namun, hasil evaluasi menunjukkan bahwa model masih memiliki keterbatasan dalam memprediksi harga dengan akurasi tinggi. Hal ini kemungkinan disebabkan oleh kompleksitas pasar mobil bekas yang dipengaruhi oleh banyak faktor lain yang tidak tercakup dalam model ini. Nilai RMSE yang tinggi (46.798.781) atau menunjukkan rata-rata nilai selisih dari prediksi harga mobil bekas \pm Rp 46.798.781,- dan R-squared yang hanya mencapai 51% menunjukkan bahwa model ini dapat dijadikan rekomendasi untuk memprediksi harga mobil bekas. Maka dari itu, perlu dilakukan pengembangan lebih lanjut dengan menambahkan variabel independen yang mayoritas tidak bersifat boolean dan memiliki korelasi dengan variabel dependen yang lebih tinggi.

DAFTAR PUSTAKA

- [1] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.
- [2] N. S. E. Kurniawati and R. G. Artaningrum, "Penerapan Strategi Pemasaran untuk Meningkatkan Penjualan Mobil Bekas di Heaven Motor," *Stud. Akuntansi, Keuangan, dan Manaj.*, vol. 4, no. 1, pp. 69–79, 2024, doi: 10.35912/sakman.v4i1.3166.
- [3] D. Miftahul Huda, G. Dwilestari, and A. Rizki Rinaldi, "Jurnal Informatika dan Rekayasa Perangkat Lunak Prediksi Harga Mobil Bekas Menggunakan Algoritma Regresi Linear Berganda," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 150–157, 2024.
- [4] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear : Studi Kasus pada Industri Otomotif di Indonesia," vol. 4, pp. 1199–1216, 2024, [Online]. Available: <https://j-innovative.org/index.php/Innovative%0APrediksi>
- [5] M. Wiharto, "Analisis Kluster Menggunakan Bahasa Pemrograman R untuk Kajian Ekologi," *J. Bionature*, vol. 14, no. 2, pp. 73–79, 2013.
- [6] Z. Fitriyah, S. Irsalina, A. R. H. K., and E. Widodo, "Analisis Faktor Yang Berpengaruh Terhadap Ipm Menggunakan Regresi Linear Berganda," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 2, no. 3, pp. 282–291, 2021, doi: 10.46306/lb.v2i3.86.