

zenius

Kampus
Merdeka
INDONESIA JAYA

Car Price Prediction

Nomor Kelompok: 3

Nama Mentor: Muhammad Ramdhan Hidayat

Nama:

- Laila Fathiyaturrahmi
- Regina Gradys Waningtyum

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



Petunjuk

- Waktu presentasi adalah 5 menit (tentatif, tergantung dari banyaknya kelompok yang mendaftarkan diri)
- Waktu tanya jawab adalah 5 menit
- Silakan menambahkan gambar/visualisasi pada slide presentasi
- Upayakan agar tetap dalam format poin-poin (ingat, ini presentasi, bukan esai)
- Jangan masukkan *code* ke dalam slide presentasi (tidak usah memasukan screenshot jupyter notebook)

1. Latar Belakang
2. Explorasi Data dan Visualisasi
3. Modelling
4. Kesimpulan

Latar Belakang

Latar Belakang Project

Sumber Data: <https://www.kaggle.com/datasets/hellbuoy/car-price-prediction>

Problem: **regression**

Tujuan:

- Menganalisis variabel yang mempengaruhi harga mobil
- Membentuk model regresi harga mobil
- Melakukan prediksi harga mobil berdasarkan variabel yang mempengaruhi

Explorasi Data dan Visualisasi

Dataset Car Price

Mobil adalah kendaraan roda empat yang digerakkan dengan tenaga mesin dan bahan bakar. Hampir setiap orang menggunakan mobil dalam beraktivitas karena keunggulan mobil seperti dapat melindungi pengguna dari panas dan hujan. Harga mobil bervariasi tergantung dengan beberapa hal seperti identitas mobil, tingkat risiko, perusahaan yang memproduksi mobil, tipe bahan bakar mobil, aspirasi mobil, jumlah pintu mobil, bentuk mobil, jenis setir, lokasi mesin mobil, jarak roda mobil, panjang mobil, lebar mobil, tinggi mobil, berat mobil tanpa penumpang atau bagasi, tipe mesin, letak silinder pada mobil, ukuran mobil, sistem bahan bakar, rasio besarnya diameter silinder mobil, volume di dalam mesin, rasio kompresi, daya, rpm maksimal mobil, jarak tempuh pada perkotaan dan jarak tempuh pada jalanan tol

Data Cleansing

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 columns):
#   Column              Non-Null Count  Dtype
---  -
0   car_ID              205 non-null    int64
1   symboling           205 non-null    object
2   CarName             205 non-null    object
3   fueltype            205 non-null    object
4   aspiration           205 non-null    object
5   doornumber          205 non-null    object
6   carbody             205 non-null    object
7   drivewheel          205 non-null    object
8   enginelocation      205 non-null    object
9   wheelbase           205 non-null    float64
10  carlength           205 non-null    float64
11  carwidth            205 non-null    float64
12  carheight           205 non-null    float64
13  curbweight          205 non-null    int64
14  enginetype          205 non-null    object
15  cylindernumber      205 non-null    object
16  enginesize          205 non-null    int64
17  fuelsystem          205 non-null    object
18  boreratio           205 non-null    float64
19  stroke              205 non-null    float64
20  compressionratio    205 non-null    float64
21  horsepower          205 non-null    int64
22  peakrpm             205 non-null    int64
23  citympg             205 non-null    int64
24  highwaympg         205 non-null    int64
25  price              205 non-null    float64
dtypes: float64(8), int64(7), object(11)
```

- Dataset Car Price terdiri dari 205 baris dan 26 kolom
- Tidak terdapat *missing value*
- Tidak terdapat *duplicate data*
- **symboling** merupakan variabel kategorik, namun terbaca sebagai integer atau numerik

Solusi : mengubah tipe data dengan menggunakan **astype**
- *Outliers* terdeteksi, namun bukan kejanggalan sehingga tidak perlu ditangani

Data Cleansing

```
['alfa-romero' 'audi' 'bmw' 'chevrolet' 'dodge' 'honda' 'isuzu' 'jaguar'  
'maxda' 'mazda' 'buick' 'mercury' 'mitsubishi' 'Nissan' 'nissan'  
'peugeot' 'plymouth' 'porsche' 'porcshe' 'renault' 'saab' 'subaru'  
'toyota' 'toyouta' 'vokswagen' 'volkswagen' 'vw' 'volvo']
```

- Ditemukan kesalahan penulisan pada brand mobil
- Solusi: dilakukan perbaikan nama brand, sehingga sebagai berikut:

```
['alfa-romero' 'audi' 'bmw' 'chevrolet' 'dodge' 'honda' 'isuzu' 'jaguar'  
'mazda' 'buick' 'mercury' 'mitsubishi' 'nissan' 'peugeot' 'plymouth'  
'porsche' 'renault' 'saab' 'subaru' 'toyota' 'volkswagen' 'volvo']
```

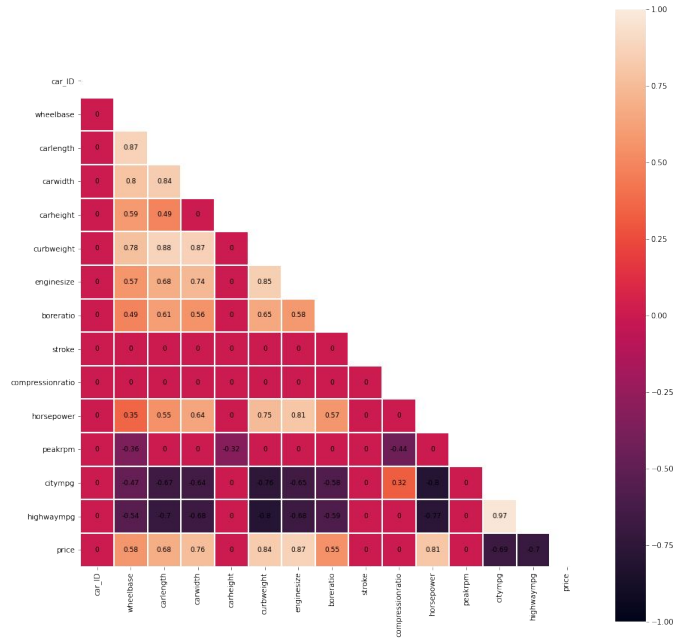
Explanatory Data Analysis

Variabel Numerik

Statistika Deskriptif

- Rata-rata harga mobil adalah \$13276.71
- Harga minimal mobil adalah \$5118
- Harga maksimal mobil adalah 45400

Explanatory Data Analysis



Korelasi Antar-variabel

- Variabel dengan nilai korelasi -0.3 sampai 0.3 dianggap **tidak** berkorelasi.

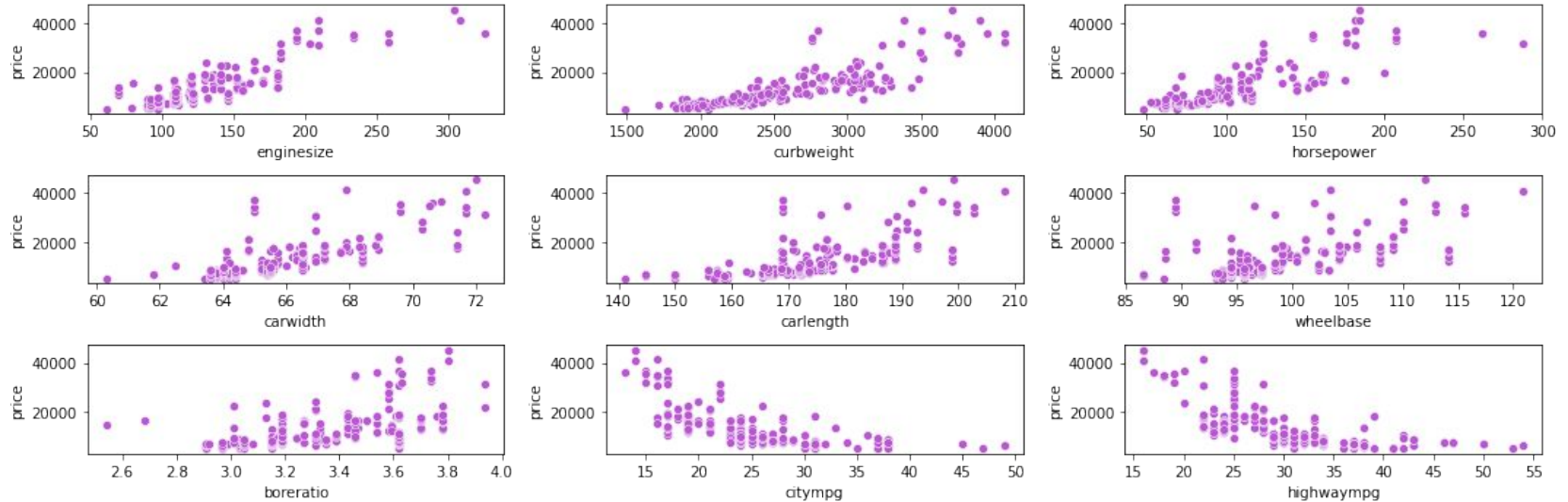
Explanatory Data Analysis

variabel	korelasi
car_ID	0
carheight	0
stroke	0
compressionratio	0
peakrpm	0
citympg	-0.686
highway	-0.698

variabel	<i>korelasi</i>
enginesize	0.874
curbweight	0.835
horsepower	0.808
carwidth	0.759
carlength	0.683
wheelbase	0.578
boreratio	0.553

“Variabel dengan korelasi diantara -0.3 sampai 0.3 kurang berpengaruh terhadap harga mobil”

Exploratory Data Analysis



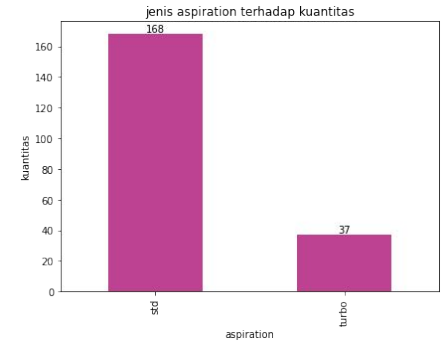
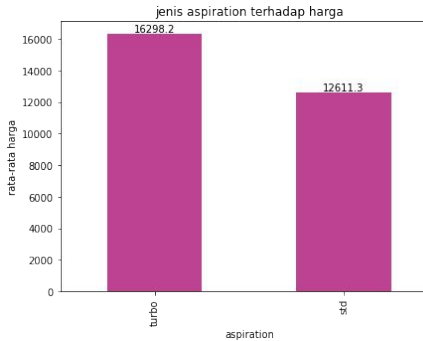
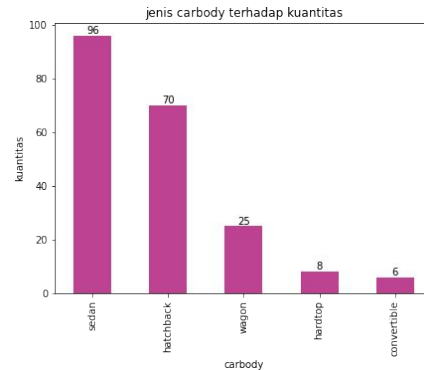
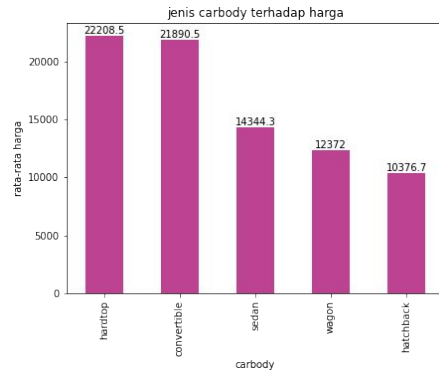
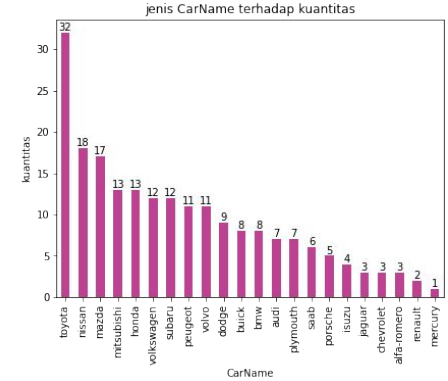
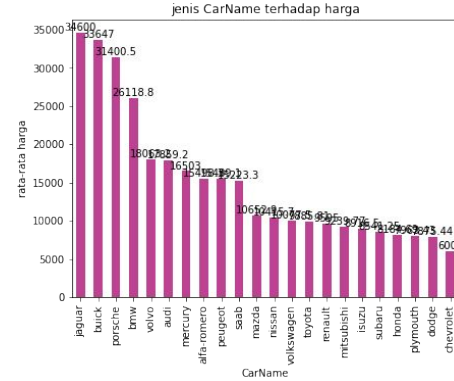
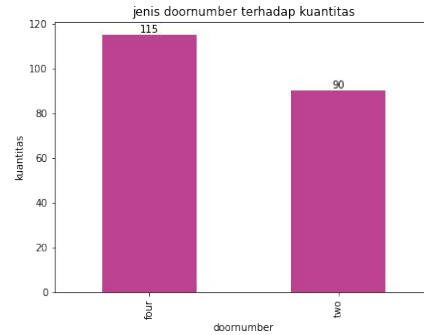
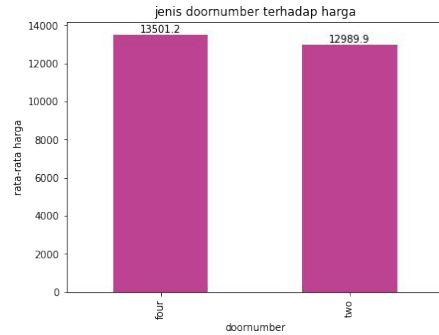
Exploratory Data Analysis

variabel	<i>p-value</i>
<i>engine</i> location	0.00
cylindernumber	0.00
fuelsystem	0.00
drivewheel	0.00
aspiration	0.00
carbody	0.00

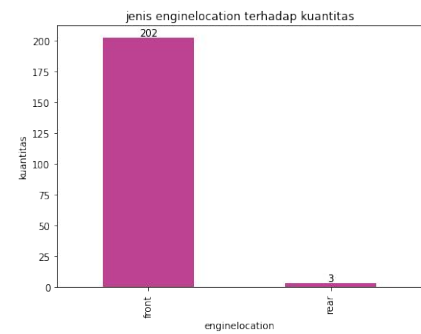
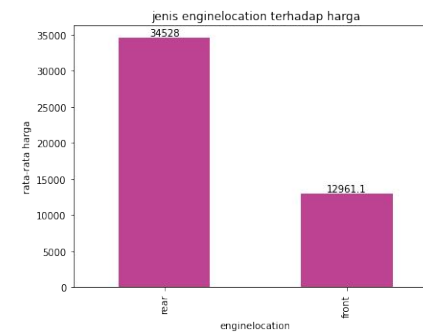
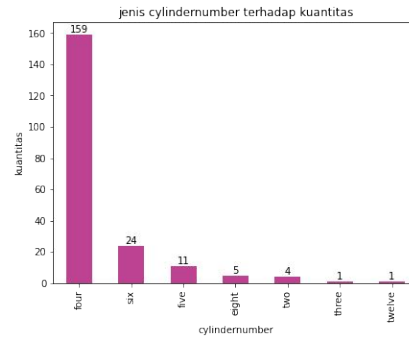
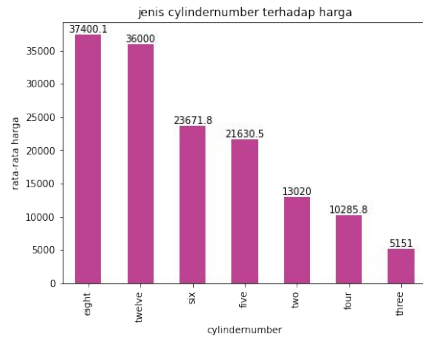
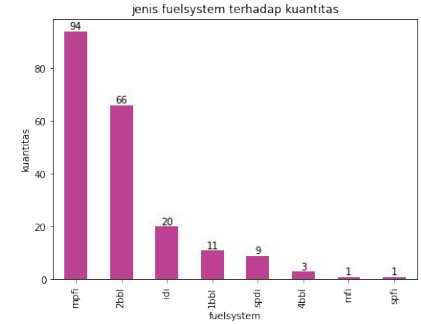
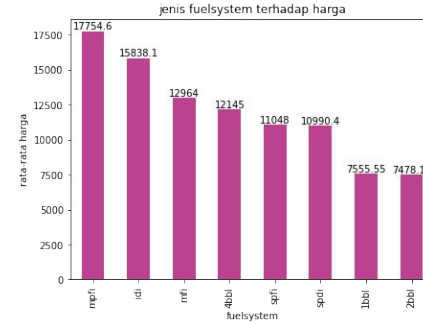
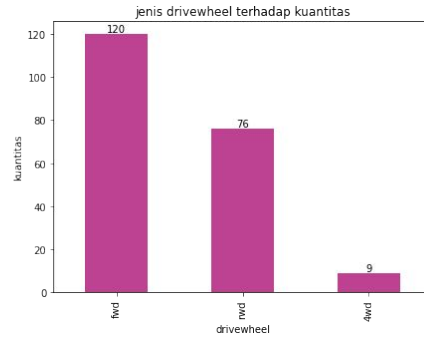
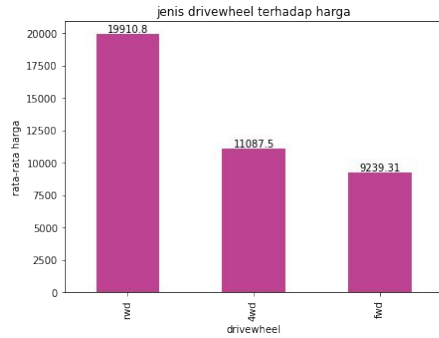
variabel	<i>p-value</i>
CarName	0.00
doornumber	0.00
symboling	0.00
fueltype	0.01
enginetype	0.03

***“Variabel
dengan
 $p\text{-value} < 0.05$
berpengaruh
signifikan
terhadap
harga mobil”***

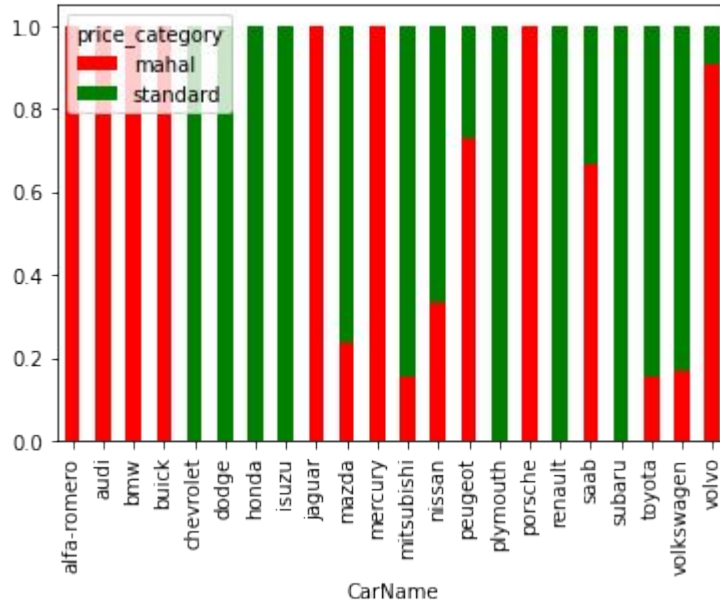
Exploratory Data Analysis



Exploratory Data Analysis



Exploratory Data Analysis



- Mengkategorikan harga mobil
 - Harga mobil < rata-rata (**standard**)
 - Harga mobil > rata-rata (**mahal**)
- Terdapat 129 mobil dengan harga standard dan 76 mobil dengan harga mahal

Modelling

Prediksi Harga Mobil

One Hot Encoding

- Mengubah variabel kategorik menjadi numerik

Train Test Split

- Mendefinisikan kolom price sebagai **y** dan kolom yang lain sebagai **x**
- Test size yang digunakan sebesar **0.3**

Model Regresi Linear

Variabel Lengkap VS Menggunakan variabel yang berpengaruh Signifikan

Model	RMSE	MSE	MAE	R-Squared
Lengkap	2474.53	6123292.1	1634.97	0.90
Signifikan	2642.95	6985171.91	1731.82	0.88

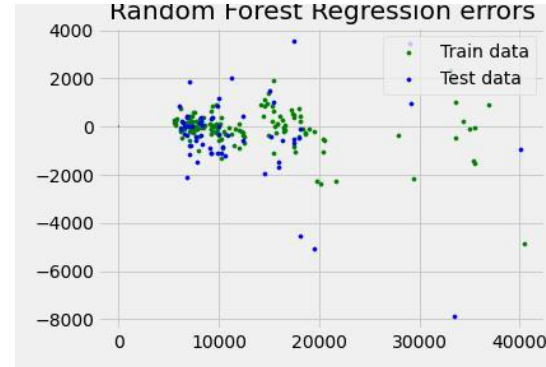
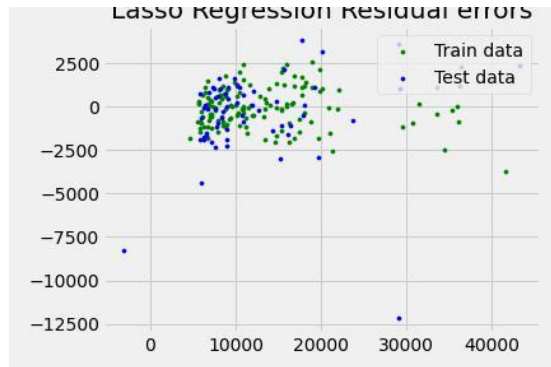
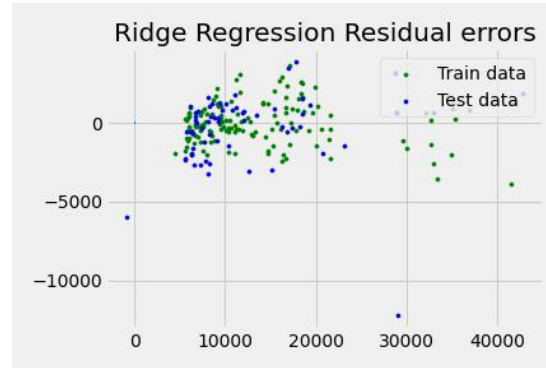
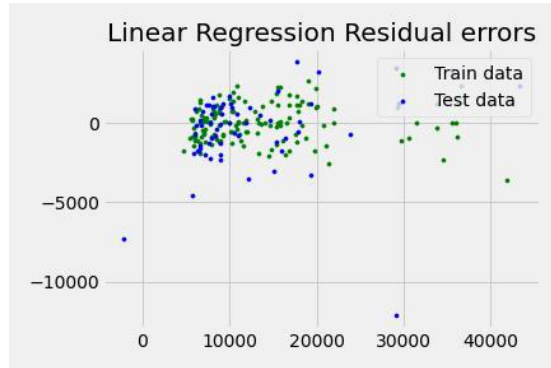
“Model Linear Regression dengan variabel lengkap lebih baik”

Evaluasi Model

Model	RMSE	MSE	MAE	R-Squared
Linear Lengkap	2474.53	6123292.1	1634.97	0.90
Linear Signifikan	2642.95	6985171.91	1731.82	0.88
Ridge	2341.63	5483228.34	1524.12	0.91
Lasso	2468.28	6092428.82	1581.11	0.90
Random Forest	1754.38	3077837.65	1088.45	0.95

“Model Random Forest lebih baik dari pada Linear Regression, Ridge Regression, dan Lasso Regression”

Evaluasi Model



“Model good fit”

Tuning

Model	RMSE	R-Squared
Linear	2474.53	0.90
Ridge	2341.63	0.91
Lasso	2468.28	0.90
Random Forest	1966.234	0.94

“
***Tuning tidak dapat
memperbaiki model***”

Model Final

“Model final yang digunakan adalah **Random Forest.**”

Dengan RMSE = 1754.38

MAE = 1088.45

R-squared = 95%

Conclusion

Conclusion

- Rata-rata harga mobil sebesar \$13276.71 dengan harga terendah sebesar \$5118 dan harga tertinggi sebesar \$45400.
- Tidak hanya brand, namun enginesize, curbweight, horsepower, carwidth, carlength, wheelbase, boreratio, citympg, highwaympg, doornumber, brand, carbody, aspiration, drivewheel, fuelsystem, cylindernumber, dan enginelocation juga berpengaruh besar terhadap harga mobil.
- Jika pembeli ingin membeli mobil dengan harga yang murah dapat memperhatikan citympg dan highway mpg yang tinggi.
- Jika pembeli ingin membeli mobil dengan harga tinggi dapat membeli mobil dengan spesifikasi enginesize, curbweight, horsepower, carwidth, carlength, wheelbase, dan boreratio tinggi, memiliki 4 pintu, brand jaguar, buick, atau porsche, dengan carbody hardtop atau convertible, memiliki aspiration turbo, drivewheel rwd, fuelsystem tipe mpfi, cylindernumber 8, dan enginelocation front.

Terima kasih!

Ada pertanyaan?

zenius



Kampus
Merdeka
INDONESIA JAYA