

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok : 3

Nama Mentor : Rauzan Sumara

Nama:

- Anis Lailatul Faadlilah
- Zahwa Nawang Sinta

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



- 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:

- Mengetahui faktor-faktor yang memengaruhi harga mobil
- Memprediksi harga mobil berdasarkan spesifikasinya
- Mengetahui model yang paling akurat dalam memprediksi harga mobil
- Membantu masyarakat dalam mengetahui perkiraan harga mobil sesuai dengan fitur yang dibutuhkan

Explorasi Data dan Visualisasi

Mobil adalah kendaraan roda empat yang digerakkan menggunakan tenaga mesin dengan bahan bakar tertentu. Mobil merupakan salah satu alat transportasi yang banyak digunakan oleh masyarakat karena dapat memudahkan seseorang dalam bepergian atau berpindah ke suatu tempat dengan nyaman.

Pada masa sekarang, mobil merupakan salah satu alat yang menjadi kebutuhan utama bagi masyarakat yang sering bepergian, khususnya yang anggota keluarganya melebihi kapasitas kendaraan rendah seperti motor. Terlebih lagi banyak perusahaan yang menciptakan produk mobil dengan fitur-fitur yang lebih bagus sehingga menetapkan harga dengan harga yang tinggi.



Oleh karena itu, kami mencoba menganalisis faktor-faktor yang dapat mempengaruhi harga mobil dengan menggunakan beberapa model, yaitu dengan **pemodelan linear regression, ridge regression, dan random forest**. Tujuannya adalah untuk membantu klien dalam menentukan fitur-fitur yang mempengaruhi harga mobil tinggi.

Data Cleansing

Tidak ada data yang missing, sehingga tidak perlu langkah penghilangan data.

```
Data columns (total 26 columns):
#  Column      Non-Null Count  Dtype
---  -
0  car_ID       205 non-null    int64
1  symboling    205 non-null    int64
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(8), object(10)
```


Data Cleansing

Tidak terdapat duplicate data

```
car_ID  symboling  CarName  fueltype  aspiration  doornumber  carbody  drivewheel  enginelocation  wheelbase  ...  
0 rows x 26 columns
```

Terdapat keganjalan pada kolom **cylinder number** dan **door number**, sehingga kita perlu mengubah data yang awalnya berupa object mejadi integer.

```
[9] df['doornumber'].head()
```

```
0    2  
1    2  
2    2  
3    4  
4    4
```

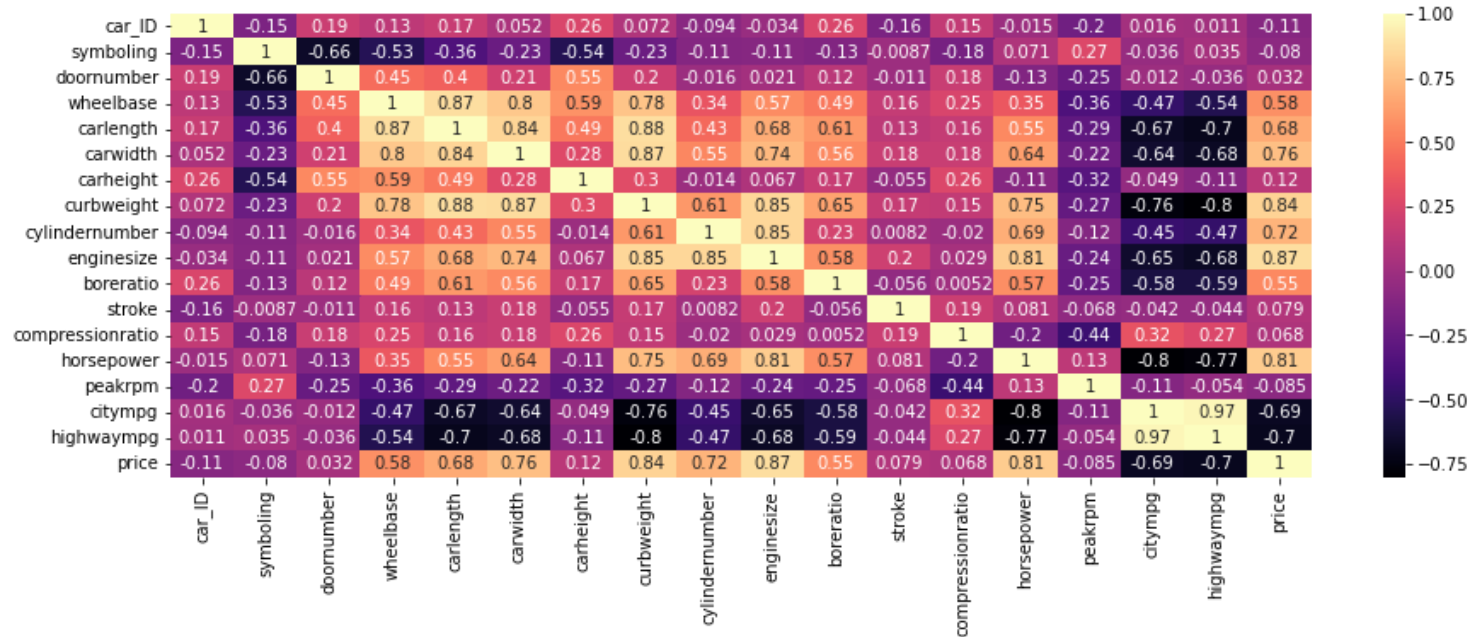
```
Name: doornumber, dtype: int64
```

```
[10] df['cylindernumber'].head()
```

```
0    4  
1    4  
2    6  
3    4  
4    5
```

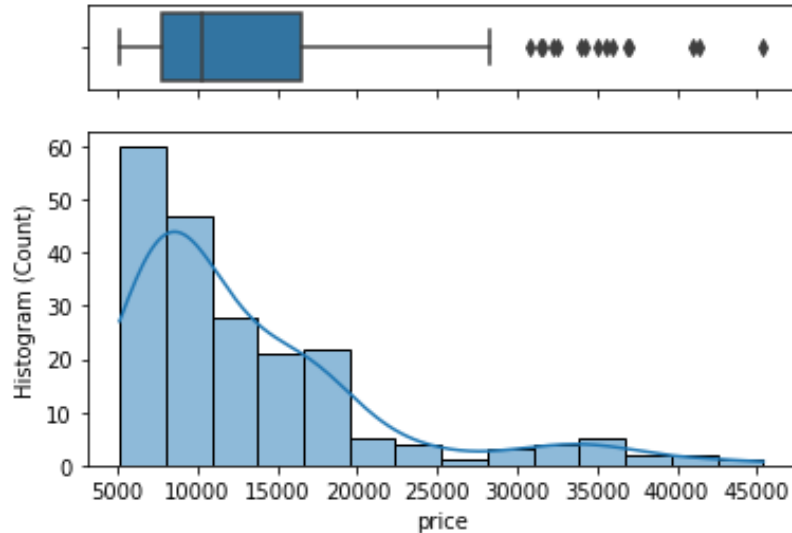
```
Name: cylindernumber, dtype: int64
```

Exploratory Data Analysis



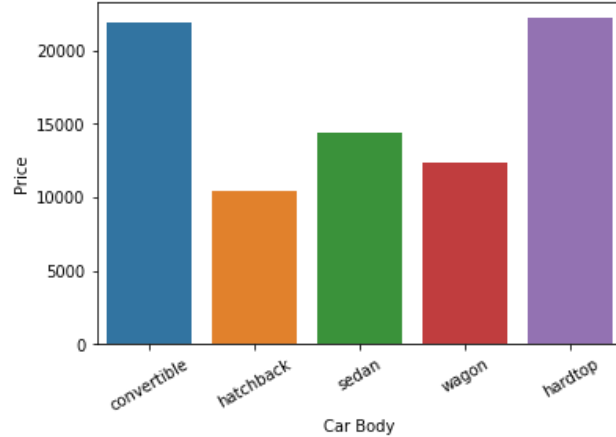
Variabel yang memiliki korelasi tinggi dengan variabel target(price) adalah wheel base, car length, car width, curb weight, cylinder number, engine size, bore ratio, horse power, city mpg, dan highway mpg.

Exploratory Data Analysis

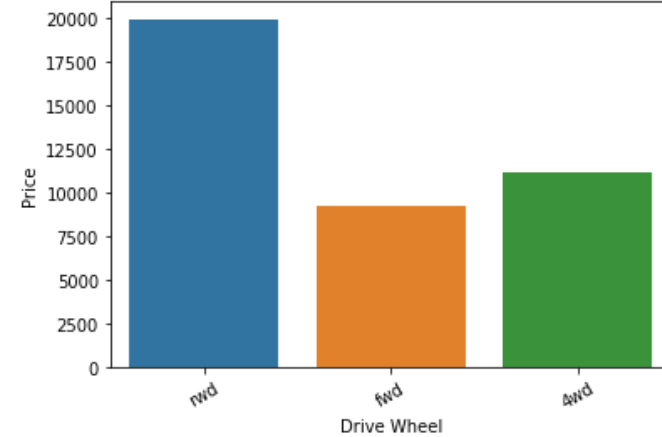


Dari gambar di samping, memperlihatkan bahwa terdapat outliers dari variabel 'price', namun karena data terlalu sedikit dan bentuk data skew right, maka outliers tidak dihapus.

Exploratory Data Analysis

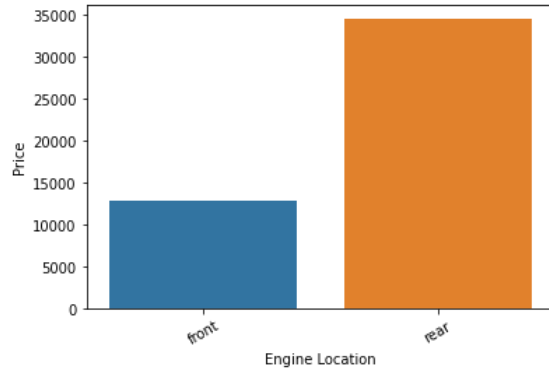


Mobil dengan tipe body convertible dan hard top mempengaruhi harga mobil lebih tinggi dibandingkan tipe body lainnya.

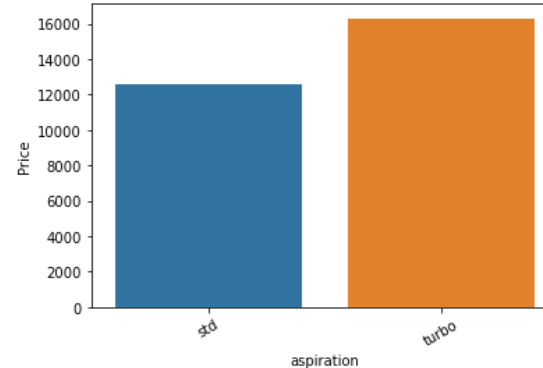


Mobil dengan tipe drive wheel rwd mempengaruhi harga mobil lebih tinggi dibandingkan tipe drive wheel fwd dan 4wd.

Exploratory Data Analysis

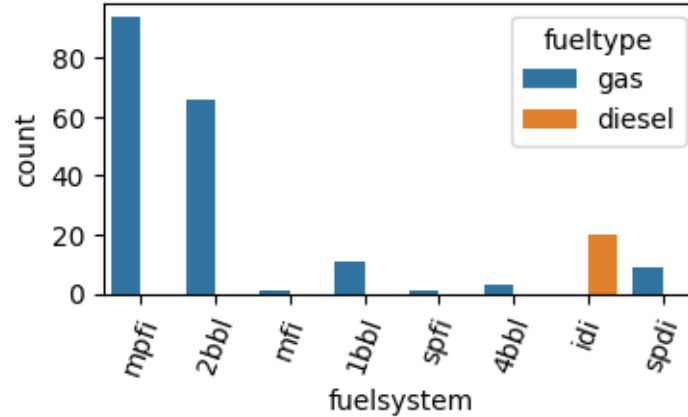


Mobil yang memiliki harga tinggi rata-rata menggunakan mesin dengan lokasi yang berada di depan dibandingkan di belakang.



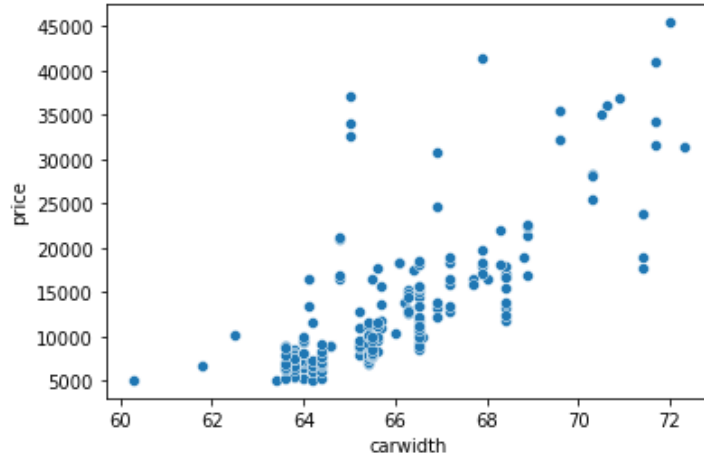
Mobil dengan aspiration tipe turbo memiliki harga yang tinggi dibandingkan dengan aspiration tipe standard.

Exploratory Data Analysis

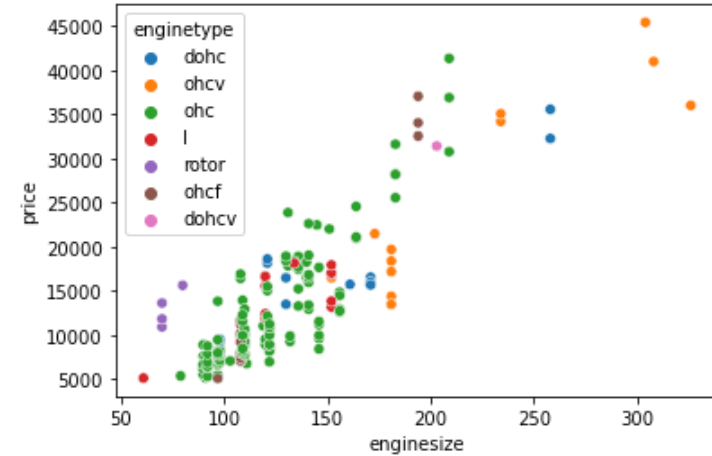


Mobil yang menggunakan tipe bahan bakar diesel hanya digunakan oleh mobil dengan sistem bahan bakar idi

Exploratory Data Analysis



Semakin besar lebar mobil, maka rata-rata harga mobil juga semakin tinggi.



Berdasarkan tiap tipe mesin, semakin besar ukuran mesin maka rata-rata harga mobil semakin tinggi.

Modelling



Model untuk Memprediksi Harga Mobil

```
#Separate into X_train, X_test, y_train, y_test (proportion 80-20)
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                    random_state=1)
```

Metrik-metrik yang digunakan :

- Mean Absolute Error
- Root Mean Squared Error
- R-Square

Linear Regression

Tujuan : membuat garis linear untuk mengoptimalkan semua titik data.

Dengan model linear regression, digunakan metrik berupa mean absolute square, root mean absolute square, dan Rsquare yang menghasilkan error sebesar :

- $MAE = 4765.509875155452$
- $RMSE = 6887.081725025669$
- $R\text{-Square} = 0.1368254872830319$

Ridge Regression

Tujuan : Membuat baris baru yang tidak terlalu fit baik dengan data train, namun tetap diharapkan memberikan hasil yang lebih baik dengan data train.

Dengan model ridge regression, digunakan metrik berupa mean absolute square, root mean absolute square, dan Rsquare yang menghasilkan error sebesar :

- $MAE = 1999.0707090604214$
- $RMSE = 3315.5761747978763$
- $R\text{-Square} = 0.7999464992096266$

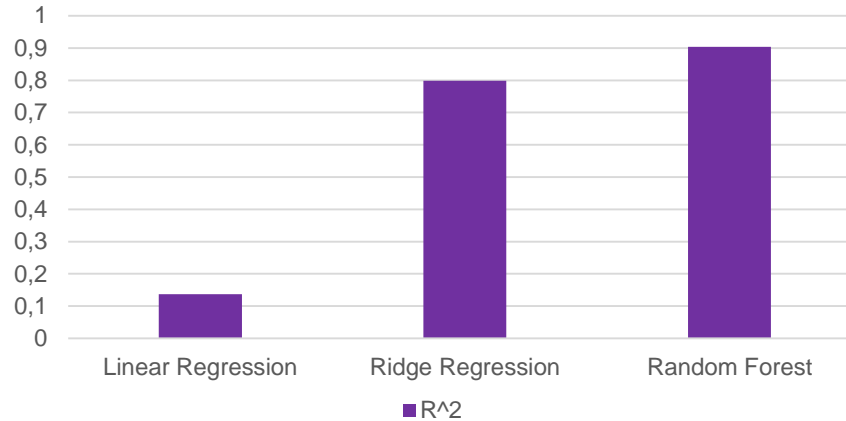
Random Forest Tuning

Tujuan : Dibuat dari himpunan bagian dari data dan hasil akhir yang didasarkan pada rata-rata atau mayoritas peringkat, dan memperbaiki masalah dari overfitting.

Dengan melakukan random forest tuning, diperoleh model baru dengan hasil error sebesar:

- RMSE = 2298.397566631523
- R-square = 0.9038656837884027
- Improvement = 1.9746972077013931%

Conclusion



Dari ketiga pemodelan yang dibuat, memperlihatkan bahwa Ridge Regression memiliki error yang lebih minimum dan nilai R-Square yang lebih tinggi daripada Linear Regression. Sedangkan, saat melakukan pemodelan random forest tuning, diperoleh pula error yang lebih minimum dari kedua model sebelumnya.

Dari random forest tuning diperoleh improvement sebesar 1.97%, sehingga model yang dipilih sebagai model final dalam melakukan prediksi harga mobil adalah pemodelan random forest.

Mobil dengan harga tinggi memiliki spesifikasi :

- highway mpg tinggi
- peak rpm tinggi
- engine size tinggi
- symboling tinggi
- door number tinggi
- wheel base tinggi
- car width tinggi
- car height tinggi
- curb weight tinggi
- city mpg rendah
- car length rendah
- cylinder number
- horse power
- bore ratio rendah
- stroke rendah
- compression ratio rendah

- fuel system dengan tipe 4bbl dan idi mempengaruhi harga mobil lebih tinggi daripada fuel system tipe mpfi
- engine type ohcv dan rotor mempengaruhi harga mobil lebih tinggi daripada type ohc
- engine location tipe rear(belakang) mempengaruhi harga mobil lebih tinggi daripada tipe front(depan)
- drive wheel tipe rwd mempengaruhi harga mobil lebih tinggi daripada tipe fwd
- car body tipe convertible dan hard top mempengaruhi harga mobil lebih tinggi daripada tipe sedan
- aspiration tipe turbo mempengaruhi harga mobil lebih tinggi daripada tipe standard
- fuel type diesel mempengaruhi harga mobil lebih tinggi daripada fuel type gas

**Terima
kasih!**
Ada pertanyaan?

zenius



**Kampus
Merdeka**
INDONESIA JAYA

