

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 7

Nama Mentor: Rachmadio Noval L

Nama:

- Zelina Bellavira Aldawiyah
- Salsabilla Rizky Putri

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 faktor-faktor yang mempengaruhi harga mobil
- Melakukan prediksi harga mobil berdasarkan faktor-faktor yang sudah dianalisis sebelumnya

Explorasi Data dan Visualisasi

Business Understanding



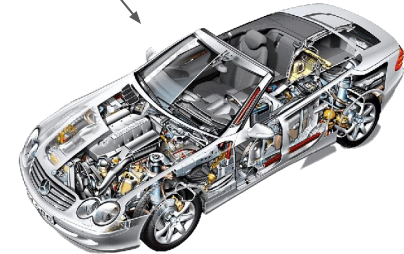
Mobil sebagai kebutuhan primer
beberapa keluarga



Perusahaan berusaha memenuhi
keinginan konsumen



Konsumen memilih faktor untuk
membeli mobil sesuai dengan
yang diinginkan



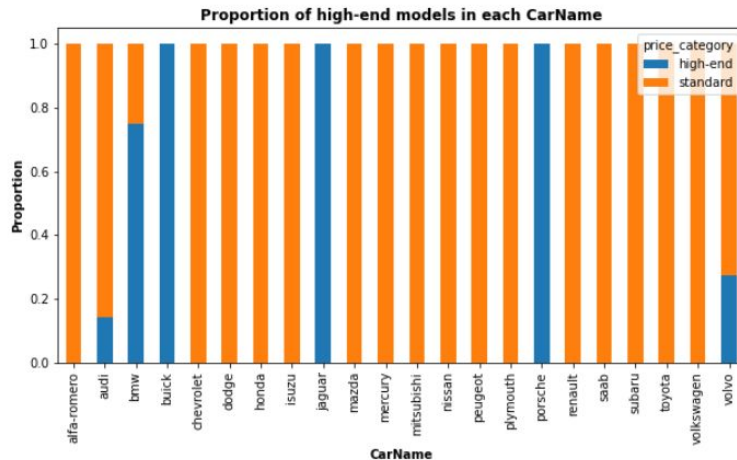
Perusahaan otomotif memodifikasi
mesin untuk keperluan pasar

Data Cleansing

- Pada dataset Car Prediction tidak terdapat missing value dan juga tidak ada value yang berisi NaN
- Dataset Car Prediction berisi 205 baris dan 26 kolom
- Melakukan drop pada kolom car_id karena tidak diperlukan untuk menganalisis harga mobil
- Mengambil value CarName hanya 1 kata terdepan saja karena kata selanjutnya merupakan type mobil
- Mengubah kolom symboling menjadi object / kategorik
- Pada kolom compression ratio, terdapat outlier namun kejanggalan tersebut dapat dijelaskan dan bukan merupakan kejanggalan yang harus diperbaiki. Sebab kejanggalan tersebut memang merupakan spesifik dari mesin mobil

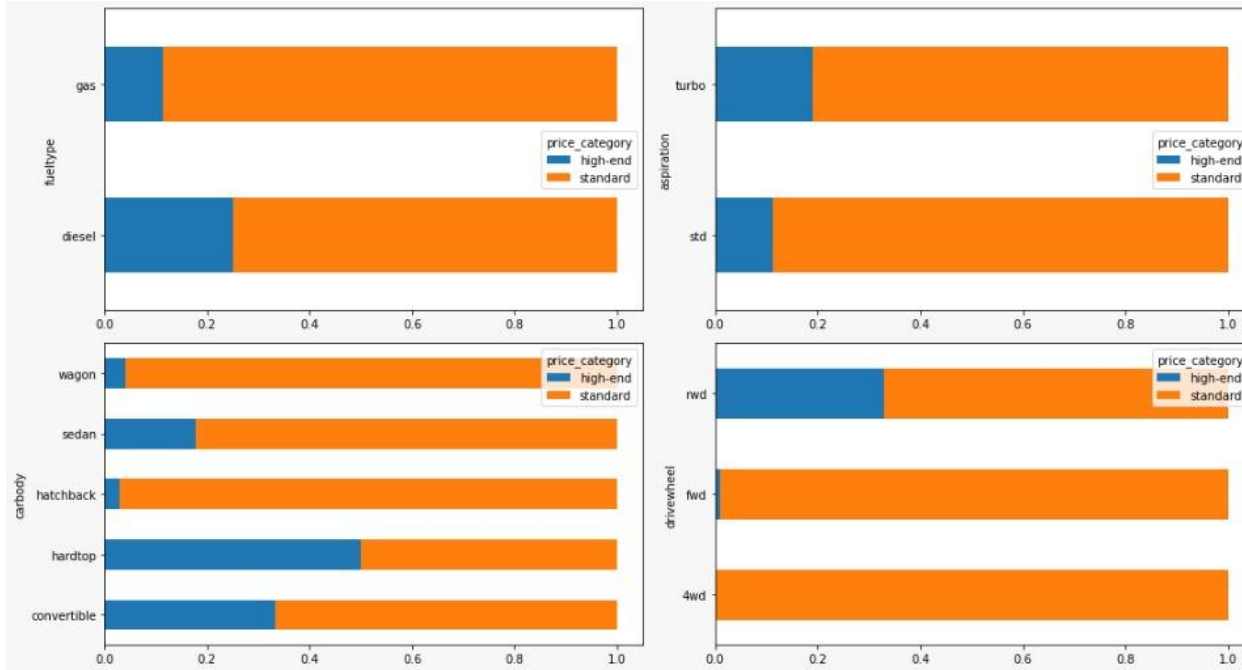
Exploratory Data Analysis

- Perbandingan harga mobil standard dan highend (>20.000) berdasarkan brand



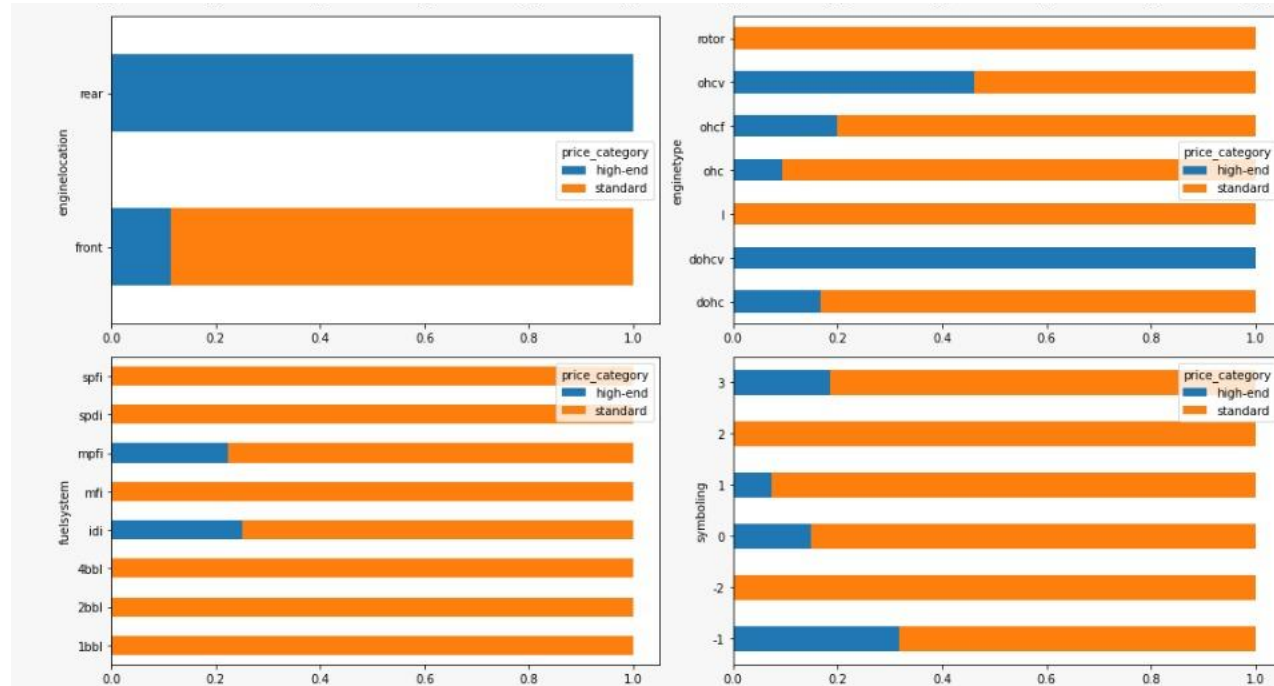
- Brand mobil yang dikatakan 'high end' ialah porsche, buick, dan jaguar.
- 80% mobil bmw merupakan mobil yang dapat dikatakan 'highend' dan 20 % lainnya dikatakan mobil yang 'standard'
- 50% mobil volvo merupakan mobil highend, dan 50% lainnya ialah standard.
- 10% mobil audi merupakan mobil highend. dan sisanya ialah mobil standard
- kategori mobil lainnya, ialah mobil yang standard atau memiliki harga di bawah 20000 dollar

Exploratory Data Analysis



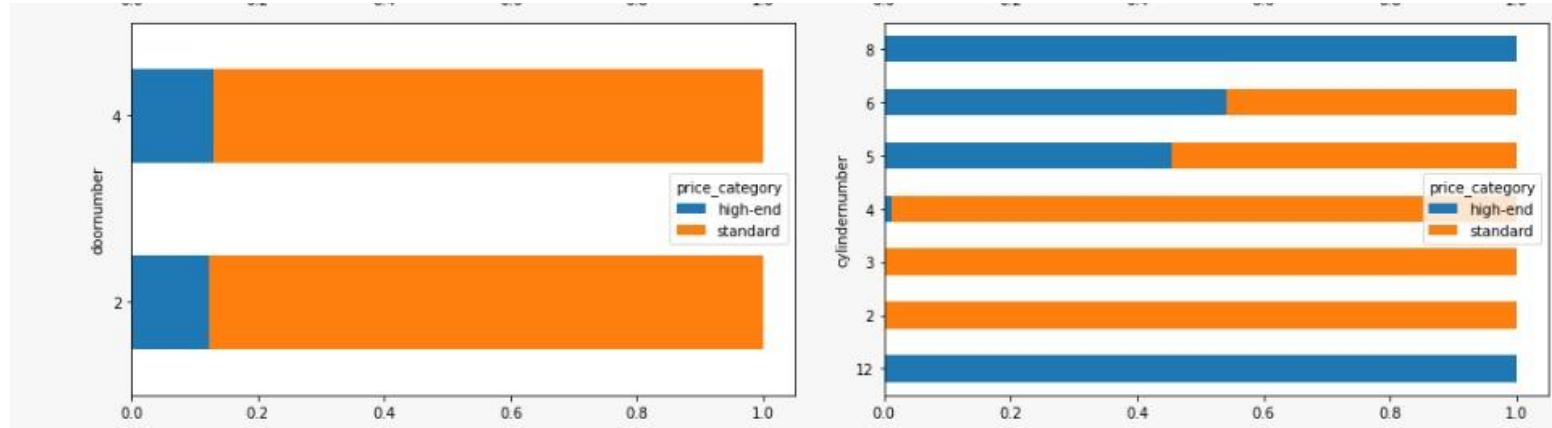
- 50% mobil yang memiliki rangka hardtop dikatakan sebagai mobil highend atau mahal, sedangkan sebagian besar mobil yang memiliki rangka wagon/hatchback merupakan mobil murah

Exploratory Data Analysis



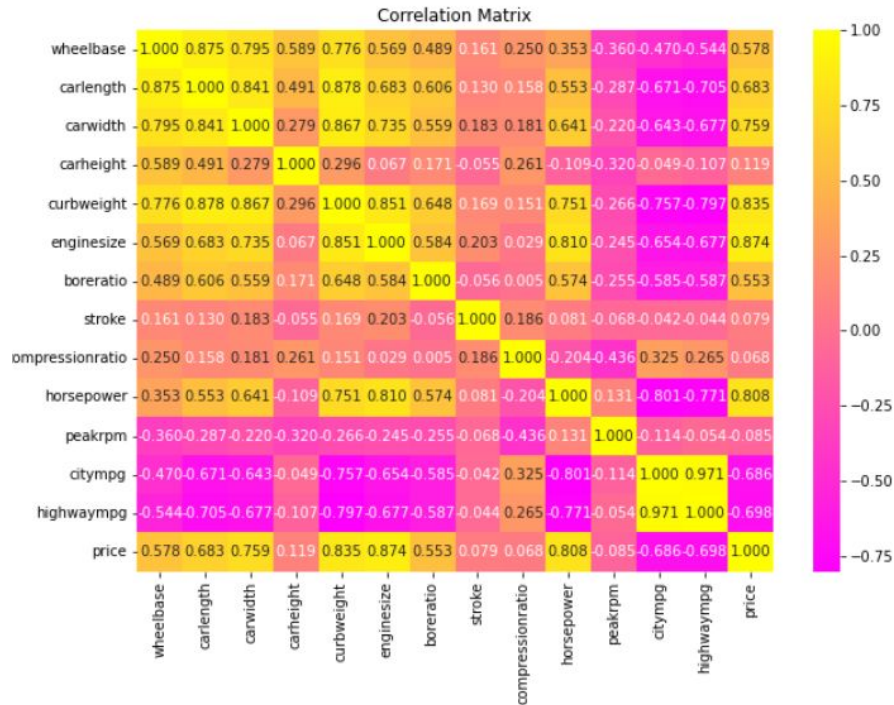
- mobil yang letak mesin berada di samping dikatakan sebagai mobil mahal
- sedangkan mobil yang letak mesin berada di depan dikatakan mobil murah

Exploratory Data Analysis



- mobil yang dikatakan mahal atau highend ialah mobil yang memiliki cylinder number sebanyak 8/12
- sedangkan mobil yang memiliki cylinder number 2/3 dikatakan sebagai mobil muran

Exploratory Data Analysis



Berdasarkan koefisien korelasi yang didapatkan pada heat map, variabel enginesize dengan price memiliki nilai korelasi tertinggi sebesar 0.874.

Variabel-variabel yang tampak dikatakan sebagai faktor yang dapat membedakan mobil 'murah' dan mobil 'mahal'

- Variabel dengan korelasi positif : wheelbase, carlength, carwidth, curbweight, enginesize, boreratio, dan horsepower
- Variabel dengan korelasi negatif : citympg dan highwaympg

Variabel lainya memiliki korelasi yang relatif sangat kecil. Sehingga diasumsikan bahwa variabel tersebut tidak dapat dikatakan sebagai faktor yang dapat membedakan mobil 'murah' dan mobil 'mahal'

Modelling



Prediksi Mobil dengan Machine Learning Regression

hal-hal yang disarankan untuk dipresentasikan:

- Untuk menjalankan model dan prediksi, membagi train set dan test set sebanyak 80 : 20. Dimana train set digunakan untuk melakukan fitting model dan test set digunakan untuk prediksi.
- Base line model
 - prediktor:** symboling, CarName, fueltype, aspiration, doornumber, carbody, drivewheel, enginelocation, enginetype, cylindernumber, fuelsystem, wheelbase, carlength, carwidth, carheight, curbweight, enginesize, boreratio, compressionratio, horsepower, peakrpm, citympg, highwaympg, stroke, price_category
 - respon :** price
- Baseline model menggunakan seluruh variabel dengan regresi linear didapatkan akurasi model sebesar 0.83 atau 83%

- Selanjutnya ialah mencoba model lain dengan menggunakan kolom-kolom terpilih untuk meningkatkan akurasi model, yaitu :

prediktor : symboling, CarName, fueltype, aspiration, doornumber, carbody, drivewheel, enginelocation, enginetype, cylindernumber, fuelsystem, wheelbase, carlength, carwidth, carheight, curbweight, enginesize, boreratio, compressionratio, horsepower, peakrpm, citympg, highwaympg, price_category

target : price

	Model	Metrics Evaluation
Baseline	Linear Regression	$R^2 = 0.83$
Evaluation Model drop kolom "stroke"	Linear Regression	$R^2 = 0.83$
	Ridge Regression	$R^2 = 0.89$
	Lasso Regression	$R^2 = 0.84$
	Random Forest Regression	$R^2 = 0.92$
	Decision Tree	$R^2 = 0.88$
	Random Forest Tuning	$R^2 = 0.90$
Final Model	Random Forest Regression	$R^2 = 0.92$

- Terdapat peningkatan yang signifikan dari akurasi model karena dilakukan drop pada kolom stroke.
- Maka model final yang digunakan untuk prediksi harga mobil ialah dengan menggunakan Random Forest Regression karena didapatkan akurasi sebesar 92%

Conclusion

- Brand mobil yang paling banyak terjual adalah brand toyota sebanyak 32 buah
- Faktor yang **paling** mempengaruhi harga suatu mobil adalah ukuran mesin. Maka semakin besar ukuran mesin, maka harga suatu mobil akan semakin tinggi. Sebaliknya apabila ukuran mesin kecil, maka harga mobil akan semakin rendah.
- Brand merupakan salah satu faktor yang mempengaruhi harga mobil

Faktor	Harga	
	Standard	High end
wheel base	Harga suatu mobil dikatakan naik apabila faktor-faktor di samping mengalami peningkatan. Semakin besar nilai faktor-faktor di samping maka harga mobil akan ikut meningkat	
carlength		
carwidth		
curbweight		
enginesize		
boreratio		
horsepower		
citympg		
highwaympg		
carbody	wagon/hatchback	hardtop
carname	chevrolet	jaguar
engine location	front	rear
cylinder number	2 atau 3	8 atau 12
enginetype	l	dohcv

Saran untuk stakeholder

- Saran yang dapat diberikan kepada perusahaan atau management otomotif yang akan memasuki pasar, dapat memilih spesifikasi mesin dan melakukan modifikasi mesin untuk menentukan harga pasar yang diinginkan.
- Pemilihan mesin ini juga dapat disesuaikan pada pasar yang ingin dituju yaitu mobil yang diproduksi nantinya berlabel highend atau yang standard.

Terima kasih!

Ada pertanyaan?

zenius



Kampus
Merdeka
INDONESIA JAYA