Final Project Presentation

Nomor Kelompok: 08
Nama Mentor: Aditya Bariq
Nama:

- Wahyu Nor Romadon
- Inayah Wijaya Adnan

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Latar Belakang

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Latar Belakang Project

Sumber Data: https://bit.ly/RegressionCarPricesPrediction
Problem: **regression**

Tujuan:

- Untuk memprediksi harga mobil berdasarkan faktor-faktor yang mempengaruhi harga mobil
- memprediksi harga mobil untuk memahami manajemen dinamika harga pasar baru
- Untuk mengetahui variabel mana yang signifikan dalam memprediksi harga mobil

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Explorasi Data dan Visualisasi

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Introduction

Mobil menjadi kendaraan paling diminati hampir diseluruh negara.

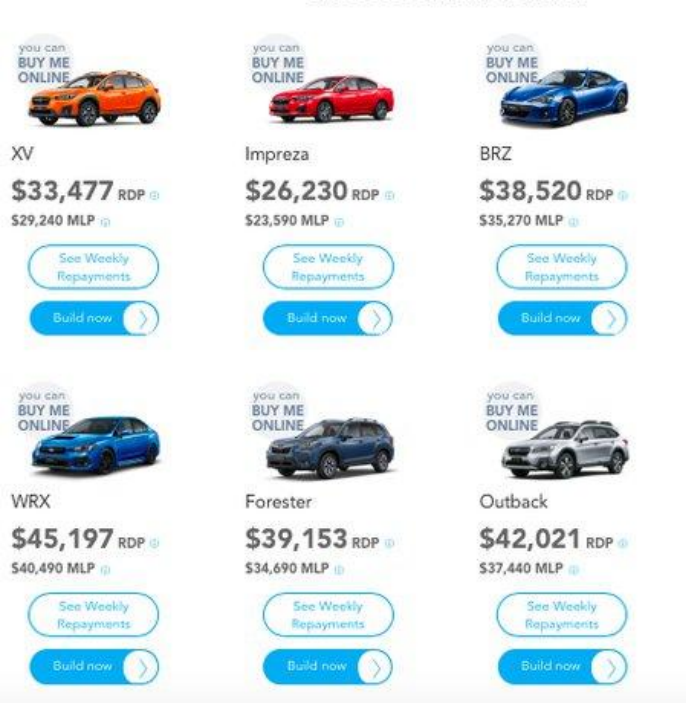
Keunggulan mobil dibandingkan motor:

- Lebih nyaman untuk perjalanan jauh
- Bisa menampung banyak penumpang
- Aman untuk bayi
- Dapat membawa barang banyak

Karena banyak keunggulannya, produk mobil sudah bervariasi dan persaingan harga mobil pun semakin ketat tiap tahunnya

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Introduction



Perusahaan harus mampu memprediksi harga mobil yang akurat untuk mampu bersaing dengan kompetitor lainnya.

Hal ini penting untuk menambah income perusahaan dan menarik pembeli sebanyak mungkin.

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Data Cleansing

- Tidak ada *missing value* pada data
- Pada kolom *CarName* ada kejanggalan yaitu adanya perbedaan tipe nama pada kolom *CarName* dan solusi dari kami yaitu menyelaraskan tipe nama *CarName* yang ada. Berikut adalah penjelasannya:

Sebelum	Sesudah
Toyouta	Toyota
Maxda	Mazda
VW	Volkswagen
vokswagen	Volkswagen
Porcshce	Porsche
nissan	Nissan

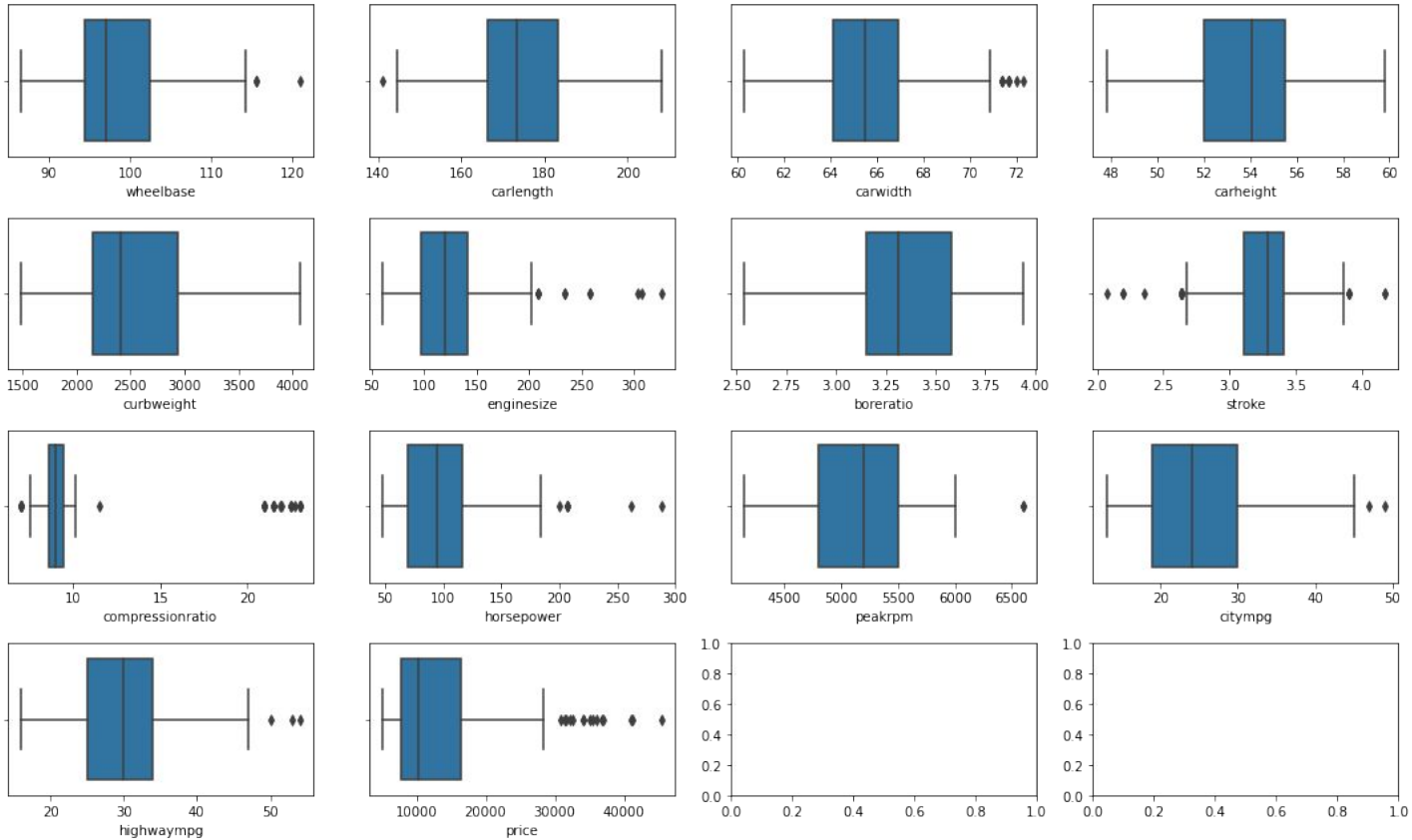
© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

BOXPLOT

- Untuk mengidentifikasi adanya outliers atau tidak
- Untuk membantu mengetahui karakteristik dari data
- Membantu kesimetrisan dari sebuah data

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

BOXPLOT



© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

BOXPLOT

Kesimpulan :

- Berdasarkan Boxplot, terlihat beberapa variabel memiliki outlier yang jaraknya cukup jauh. Namun, hal tersebut merupakan hal yang sah. Kejanggalan tersebut dapat dijelaskan dan bukan merupakan kejanggalan yang harus diperbaiki. Sebab kejanggalan tersebut memang merupakan spesifik dari mesin mobilnya

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

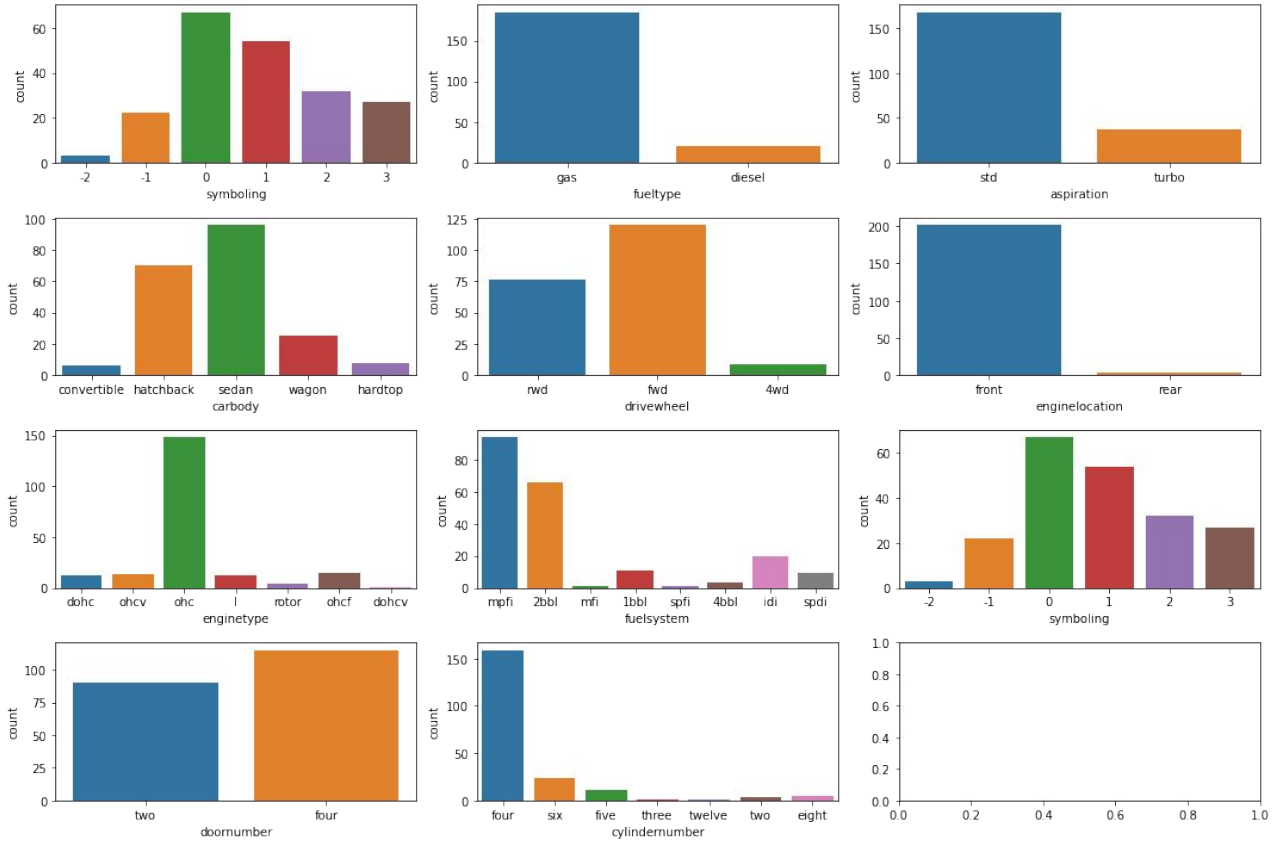
Exploratory Data Analysis

Kita bagi menjadi dua yaitu :

1. Data Kategori
2. Data Numerik

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Exploratory Data Analysis - Data Kategori



© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Exploratory Data Analysis - Data Kategori

Kolom Kategori	P-Value
enginelocation	0.000000e+00
cylindernumber	7.388765e-09
fuelsystem	2.062501e-04
drivewheel	2.174127e-04
aspiration	2.833995e-04
carbody	5.629716e-03
CarName	4.661177e-02
doornumber	4.760354e-02
symboling	6.744775e-02
fueltype	1.973501e-01
enginetype	3.512990e-01

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Exploratory Data Analysis - Data Kategori

P-value

Hipotesis :

- H0 = Harga mobil murah
- H1 = Harga mobil mahal

Dapat disimpulkan dari tabel sebelumnya bahwa kolom enginelocation, cylindernumber, fuelsystem, drivewheel, aspiration, carbody, carname, doornumber memiliki nilai p-value<0,05 maka tolak H0.

Hal itu menunjukan bahwasannya kolom tersebut mempengaruhi harga mobil menjadi mahal.

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Exploratory Data Analysis - Data Kategori

CarName	symboling	cylindernumber	
Nissan	0	four	9249.000000
		six	13799.000000
	1	four	7115.666667
		six	18399.000000
	2	four	8249.000000
	3	six	18449.000000
alfa-romero	1	six	16500.000000
	3	four	14997.500000
audi	0	five	17859.167000
	1	five	20168.333333
	2	five	16350.000000
		four	13950.000000
bmw	0	four	16925.000000
		six	30206.000000
	1	six	24565.000000
	2	four	16430.000000
buick	-1	eight	34184.000000
		five	28466.666667
	0	eight	40960.000000
		five	28176.000000
	1	eight	45400.000000
	3	eight	35056.000000
chevrolet	0	four	6575.000000
	1	four	6295.000000
	2	three	5151.000000
dodge	-1	four	8921.000000
	1	four	6999.142857
	3	four	12964.000000

Dapat disimpulkan bahwasannya mobil yang memiliki peluang lebih aman dan silinder mobil lebih banyak jauh lebih mahal.

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Exploratory Data Analysis - Data Numerik

EDA untuk Data numerik menggunakan beberapa metode antara lain :

- Statistik Deskriptif
- Heat Map
- Scatter Plot

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

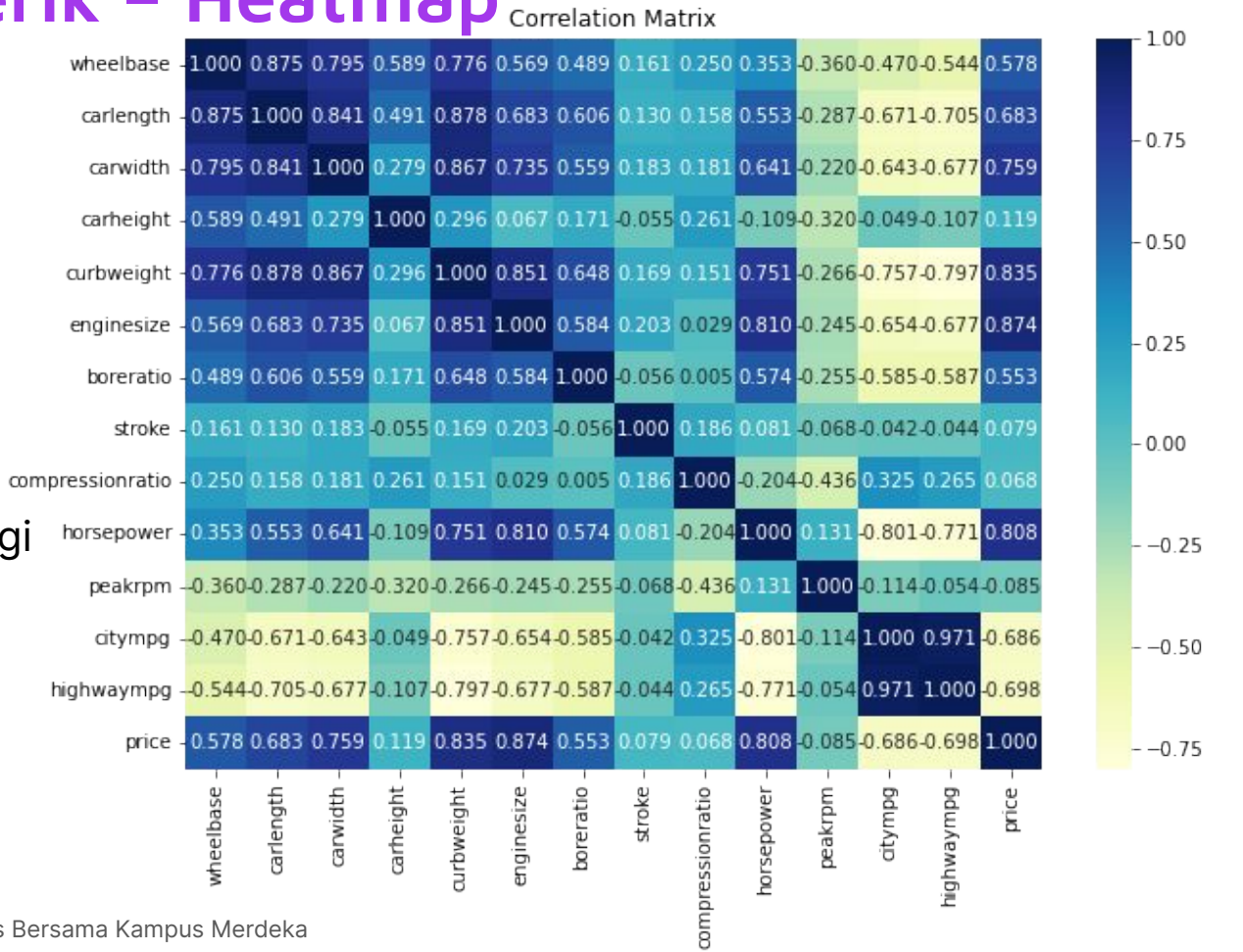
EDA- Data Numerik – Statistik Deskriptif

	count	mean	std	min	25%	50%	75%	max
wheelbase	205.0	98.756585	6.021776	86.60	94.50	97.00	102.40	120.90
carlength	205.0	174.049268	12.337289	141.10	166.30	173.20	183.10	208.10
carwidth	205.0	65.907805	2.145204	60.30	64.10	65.50	66.90	72.30
carheight	205.0	53.724878	2.443522	47.80	52.00	54.10	55.50	59.80
curbweight	205.0	2555.565854	520.680204	1488.00	2145.00	2414.00	2935.00	4066.00
enginesize	205.0	126.907317	41.642693	61.00	97.00	120.00	141.00	326.00
boreratio	205.0	3.329756	0.270844	2.54	3.15	3.31	3.58	3.94
stroke	205.0	3.255415	0.313597	2.07	3.11	3.29	3.41	4.17
compressionratio	205.0	10.142537	3.972040	7.00	8.60	9.00	9.40	23.00
horsepower	205.0	104.117073	39.544167	48.00	70.00	95.00	116.00	288.00
peakrpm	205.0	5125.121951	476.985643	4150.00	4800.00	5200.00	5500.00	6600.00
citympg	205.0	25.219512	6.542142	13.00	19.00	24.00	30.00	49.00
highwaympg	205.0	30.751220	6.886443	16.00	25.00	30.00	34.00	54.00
price	205.0	13276.710571	7988.852332	5118.00	7788.00	10295.00	16503.00	45400.00

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

EDA- Data Numerik – Heatmap

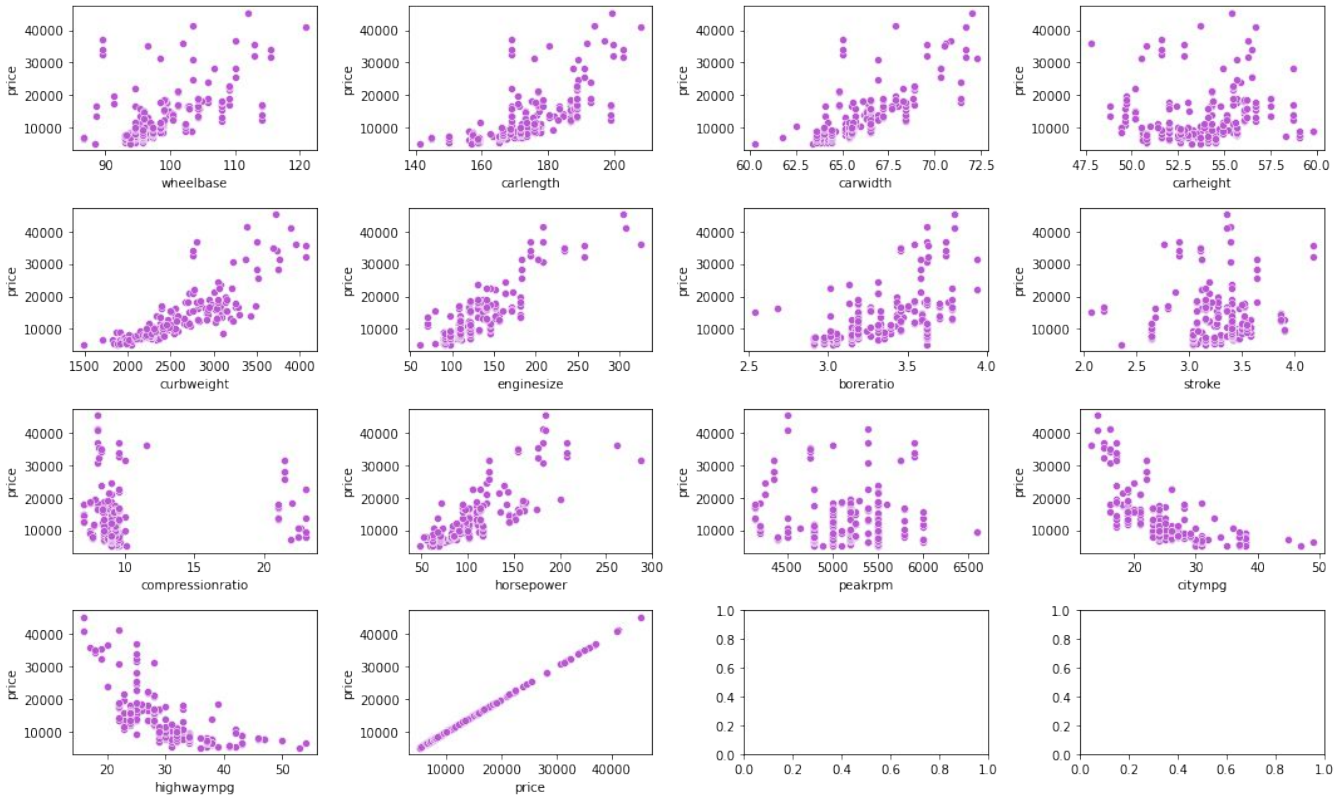
Dapat disimpulkan bahwa Berdasarkan koefisien korelasi yang didapatkan pada Heat Map, variabel enginesize dengan price memiliki nilai korelasi tertinggi sebesar 0.874



© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

EDA- Data Numerik – Scatter Plot

Pada scatter plot variabel enginesize yang memiliki visualisasi sebaran data yang membentuk pola garis lurus dengan korelasi positif.
Dimana ketika enginesize meningkat, maka harga mobil juga akan meningkat.



© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

EDA- Data Numerik – Scatter Plot

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

Variabel dengan korelasi positif	Variabel dengan korelasi negatif
Wheelbase, carlength, carwidth, curbweight, enginesize, boreratio, dan horsepower	Citympg dan highwaympg

Untuk variabel lainnya, dapat dilihat memiliki korelasi yang relatif sangat kecil dan pada scatter plot sebaran data yang ada tidak menunjukkan suatu pola atau dapat dikatakan memiliki sebaran acak. Sehingga diasumsikan variabel tidak dapat dikatakan sebagai faktor yang dapat membedakan mobil ‘murah’ dan mobil ‘mahal’

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Modelling

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Metode train test split

Menentukan kolom yang akan menjadi variabel dependen dan menjadi variabel independen

y = price

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

Menentukan dan membagi data-data menjadi data training dan data test. Data training sebanyak 70% dari jumlah data dan data test sebanyak 30%

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Evaluasi Metrik

MAE	1575.8689932307564
MSE	5558586.85531815
RMSE	2357.665552048922
R2	0.9079256289634281

MAE (rata-rata error) : dapat diinterpretasikan bahwa prediksi Price akan error +1575.8689932307564 atau -1575.8689932307564 dari harga aslinya.

RMSE : dapat diinterpretasikan bahwa prediksi Price akan error +2357.665552048922 atau -2357.665552048922 dari harga aslinya.

R2 : diketahui bahwa akurasi model regresi yang terbentuk ialah sebesar 70%. Hal itu menandakan bahwa model regresi yang tebentuk sudah cukup kuat. Selain itu bisa dijelaskan bahwa variabel independen dapat mempengaruhi variabel dependen (Price) sebesar 70% dan 30% lainnya dipengaruhi variabel lain di luar model.

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

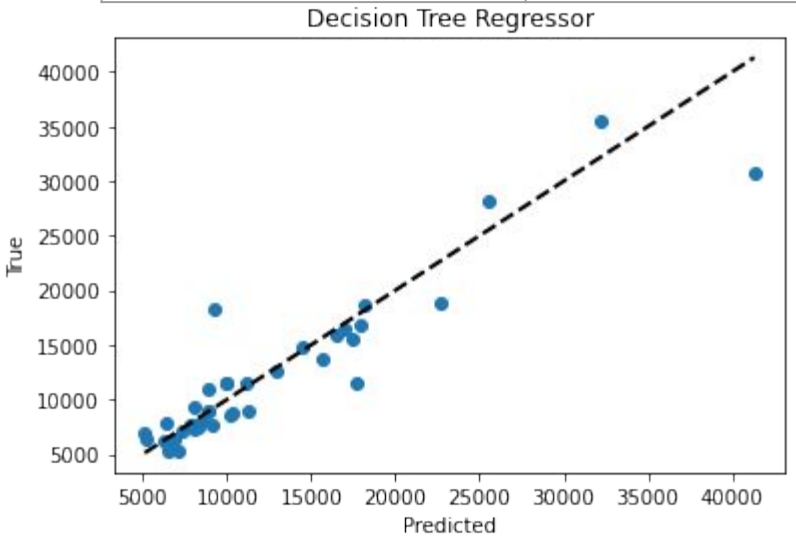
Model Pertama

Model	RMSE	MSE	MAE	R2
LinearRegression	2351.68	5530376.55	1548.67	0.91
LassoRegression	2428.59	5898063.30	1535.67	0.90
RidgeRegression	2367.55	5605312.34	1588.88	0.91

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Decision Tree

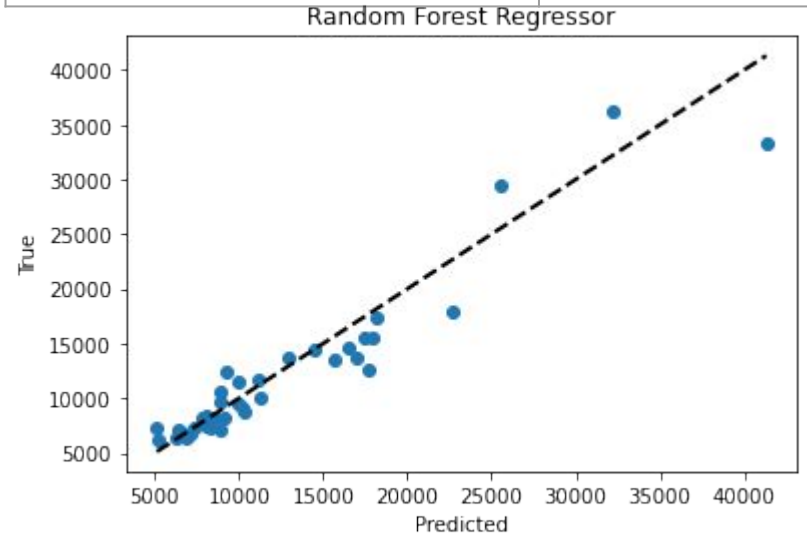
Model	RMSE	MSE	MAE	R2
Decision Tree Regressor	2543.60	6469907.88	1567.63	0.88



© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Random Forest

Model	RMSE	MSE	MAE	R2
Random Forest Regressor	2035.60	4143670.79	1355.98	0.93



Berdasarkan pemodelan yang ada, model terbaik adalah **Random Forest** dengan R2-Score sebesar 0.93

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Random Forest Tuning

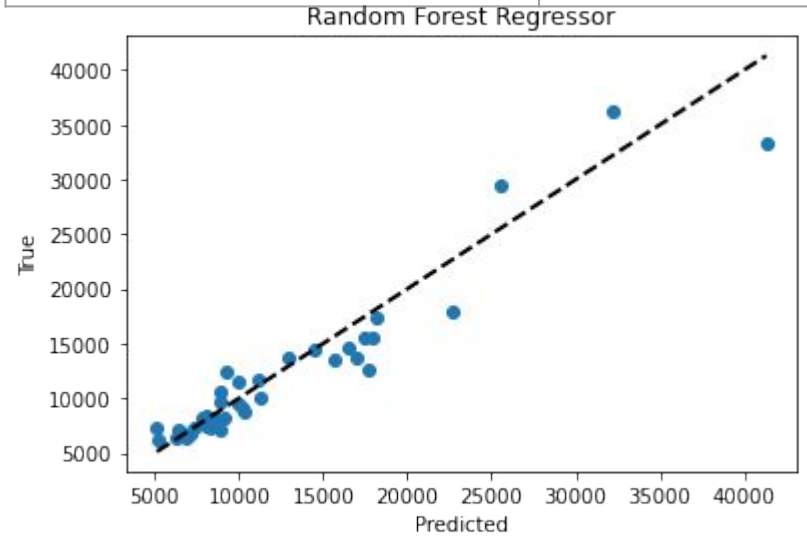
Model	RMSE	R2
Random Forest Regressor	2446.202756393957	0.891103721626432

Hyperparameter tuning pada Random Forest tidak membuat model menjadi lebih baik

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Model Final

Model	RMSE	MSE	MAE	R2
Random Forest Regressor	2035.60	4143670.79	1355.98	0.93



Berdasarkan pemodelan yang ada, model terbaik adalah **Random Forest** dengan R2-Score sebesar 0.93

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Kolom yang akan menjadi variabel dependen dan menjadi variabel independen

y = price

X = symboling, CarName, fueltype, aspiration, doornumber, carbody, drivewheel, enginelocation, wheelbase, carlength, carwidth, carheight, curbweight, enginetype, cylindernumber, enginesize, fuelsystem, boreratio, stroke, compressionratio, horsepower, peakrpm, citympg, highwaympg.

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Conclusion

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Kesimpulan

Berdasarkan apa yang telah dipaparkan maka dapat disimpulkan bahwa

1. Kebanyakan konsumen membeli mobil yang memiliki keamanan standart
2. Tingginya harga mobil diprediksi dikarenakan beberapa komponen mobil antara lain Wheelbase, carlength, carwidth, curbweight, enginesize, boreratio, horsepower, enginelocation, cylindernumber, fuelsystem, drivewheel, aspiration, carbody, carname, dan doornumber.
3. mobil yang memiliki peluang lebih aman dan silinder mobil lebih banyak jauh lebih mahal.
4. Mobil dengan keamanan tinggi dan mesin yang handal memiliki harga yang tinggi sedangkan mobil dengan keamanan standart dan mesin yang standart memiliki harga yang ekonomis, sehingga pembeli dapat mempertimbangkan pembelian mobil sesuai kebutuhan dan budget yang dimilikinya.

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Saran

Stakeholder sebaiknya mempertimbangkan target pasar konsumennya. Dengan pertimbangan sebagai berikut.

a. Mobil yang memiliki harga yang mahal rentan terjadi penurunan pembelian konsumen
b. Produk mobil akan mengalami peningkatan jumlah pembelian apabila stakeholder menambah produksi mobil dengan mesin standart (jumlah cylinder lebih sedikit) dan keamanan standart.

© 2022 Program Studi Independen Bersertifikat Zenius Bersama Kampus Merdeka

Terima kasih!

Ada pertanyaan?