

Machine Learning Models

AUTOMOBILE DATASET

André Areosa | Gonçalo Pereira | João Alves | João Simões | Tiago Cardoso

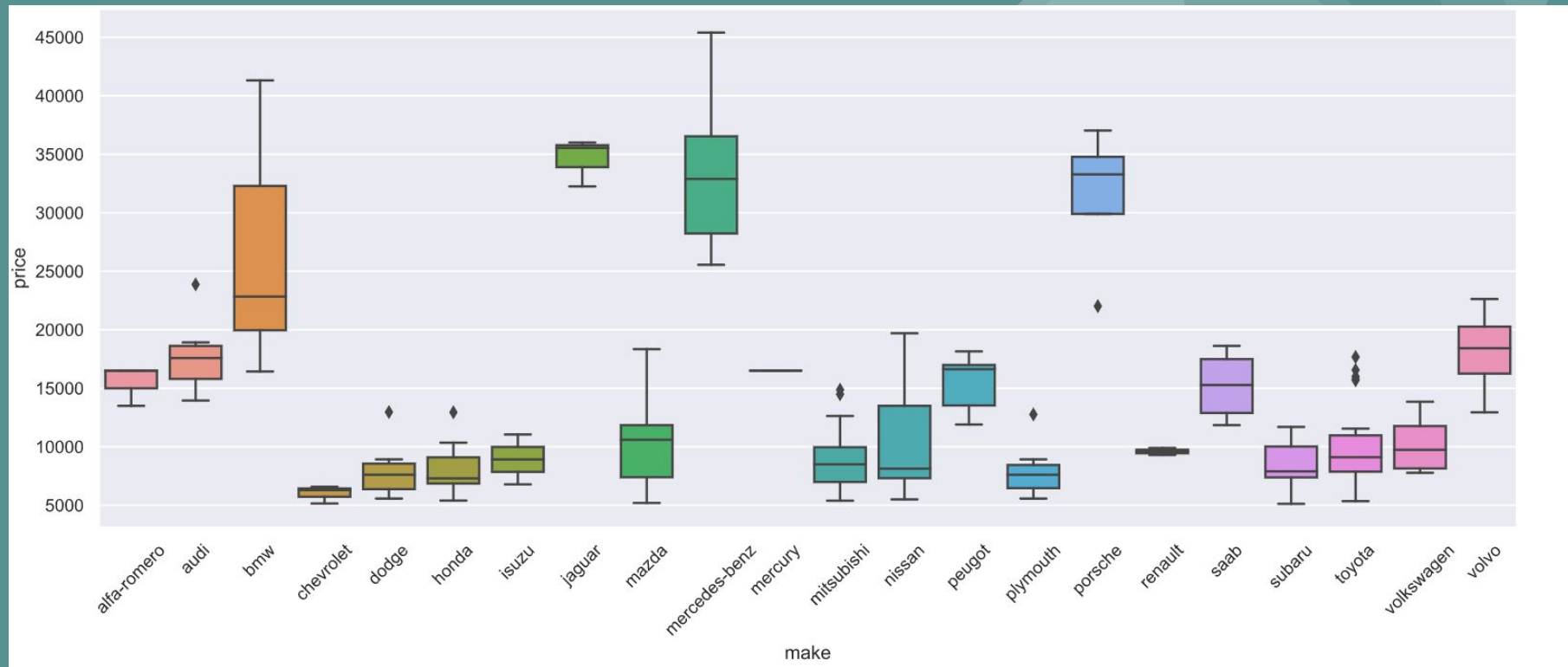
18/06/2021



Business problem

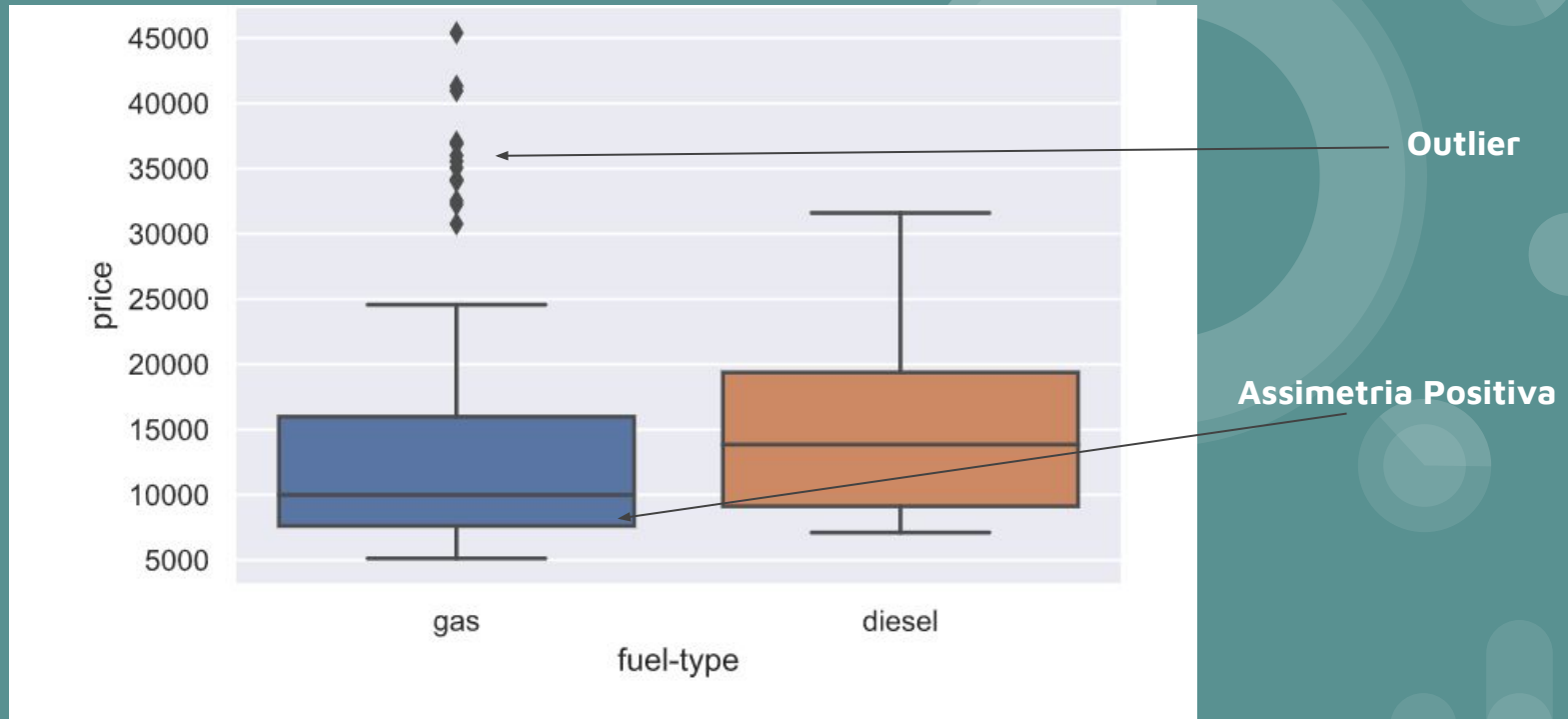
Prever o preços dos carros em função
das suas características

Data understanding



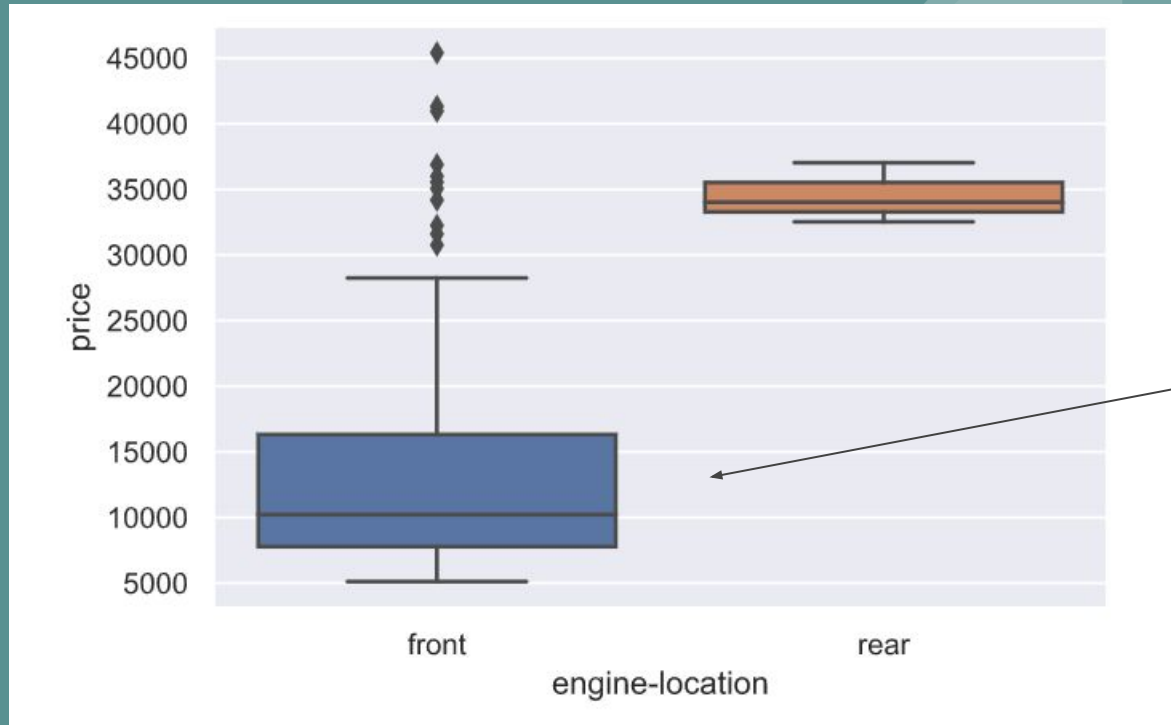
Neste gráfico pretendemos observar a dispersão entre marcas e o intervalo de preço por Marca “Symboling”

Data understanding



Neste gráfico pretendemos observar a dispersão entre carros a diesel e carros a gasolina

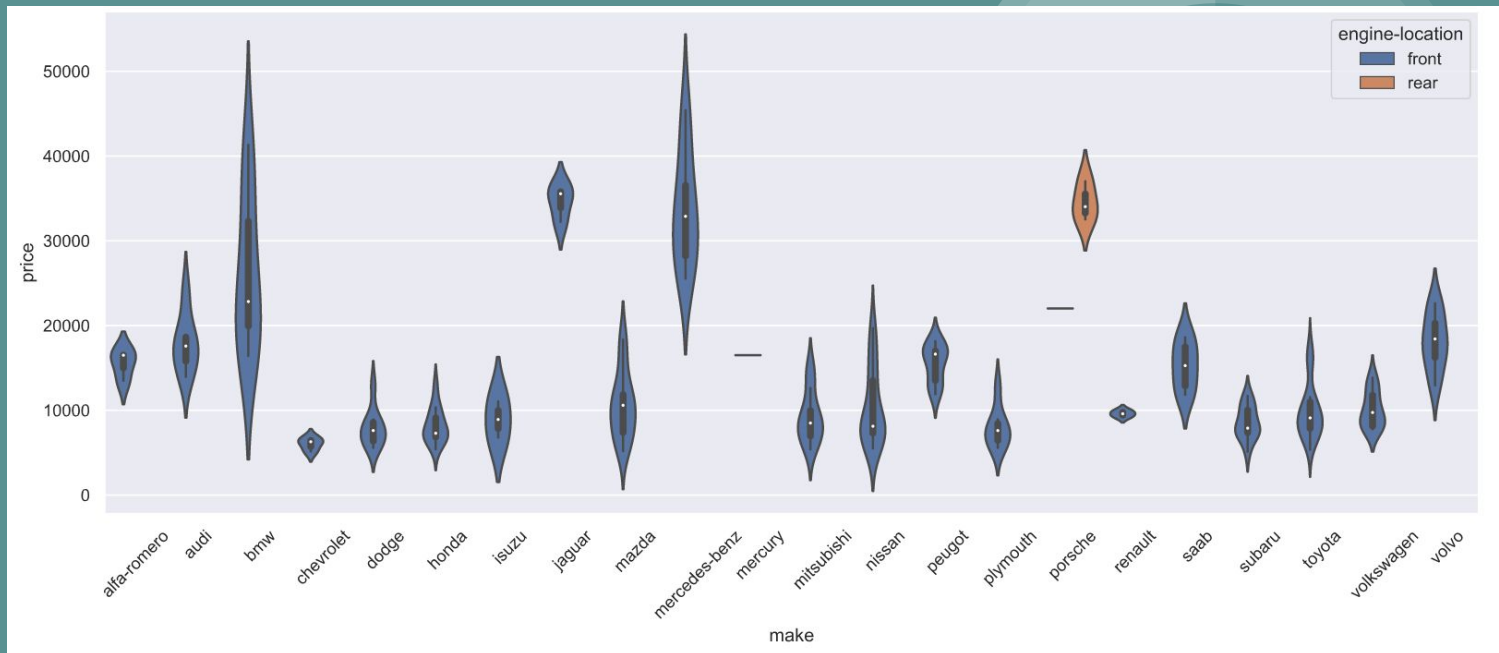
Data understanding



Assimetria Positiva

Nesta visualização é possível verificar qual o intervalo de preços dos carros com posição de motor dianteira e traseira

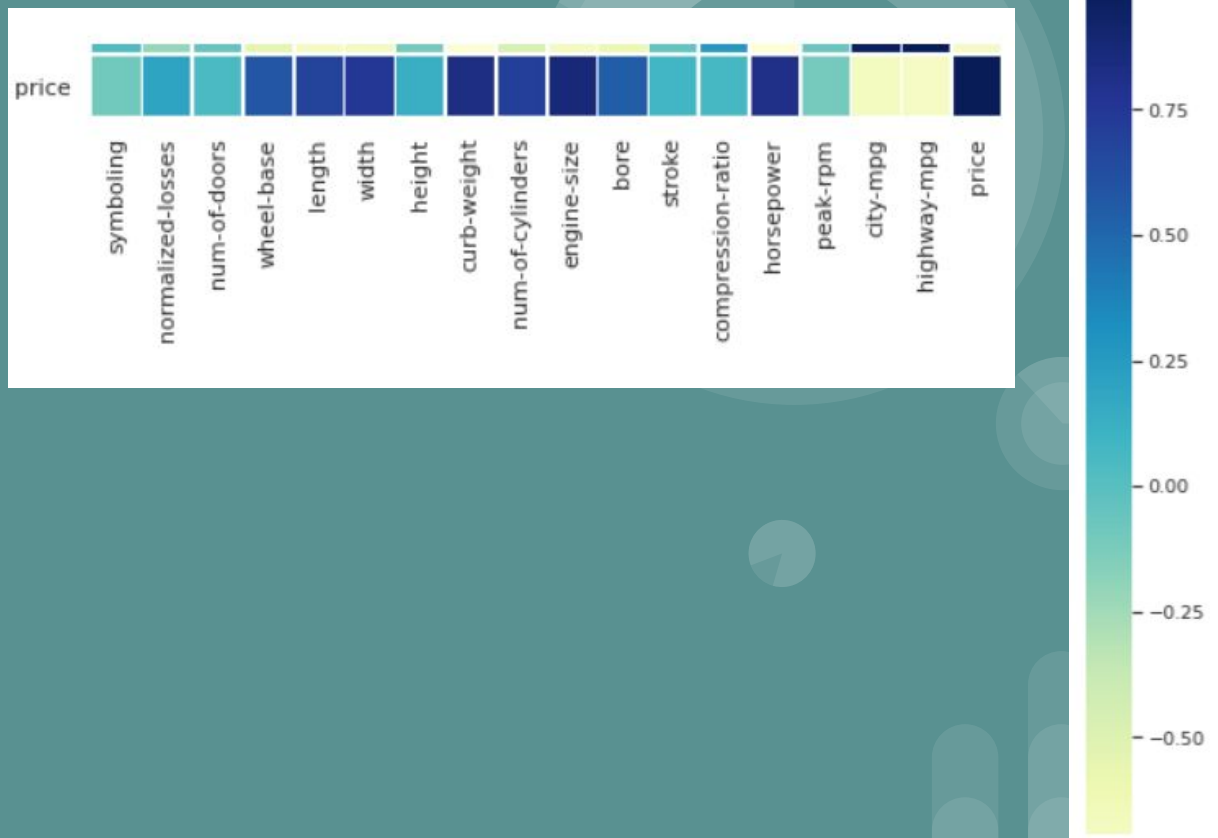
Data understanding



Como o valor de tração traseira era de pouca relevância, verificamos que só a marca porque é que apresenta carros de tração traseira

Data understanding

price	1.000000
engine-size	0.872335
curb-weight	0.834415
horsepower	0.810533
width	0.751265
num-of-cylinders	0.708645
length	0.690628
wheel-base	0.584642
bore	0.543436
normalized-losses	0.203254
height	0.135486
stroke	0.082310
compression-ratio	0.071107
num-of-doors	0.046532
symboling	-0.082391
peak-rpm	-0.101649
city-mpg	-0.686571
highway-mpg	-0.704692



Data preparation

- Conversão do números de cilindros e número de portas em valores numéricos (números “escritos por extenso” -> int)
- Conversão Float->Int de “Bore”, “Stroke” e “Compression Ratio” (através de x10)
- Correção de data types
- Remove/replace strange numbers e/ou valores nulos
 - Eliminação de linhas sem “Price”;
 - Input de valores em falta de “Nr of doors” através do tipo de veículo;
 - Eliminação das linhas sem “peak-rpm”
 - Eliminação de coluna “Stroke” pois para “ Engine-type ”=“ Rotor ” não existe
 - Impute da mediana para “Bore” = N/A

Resultado Modelos Testados

Modelos testados :

- Regressão Linear
- GridSearch
 - LGBM
 - RandomForest

Linear Regression

RMSE : 2199
R2 : 0,927

Random Forest Regressor

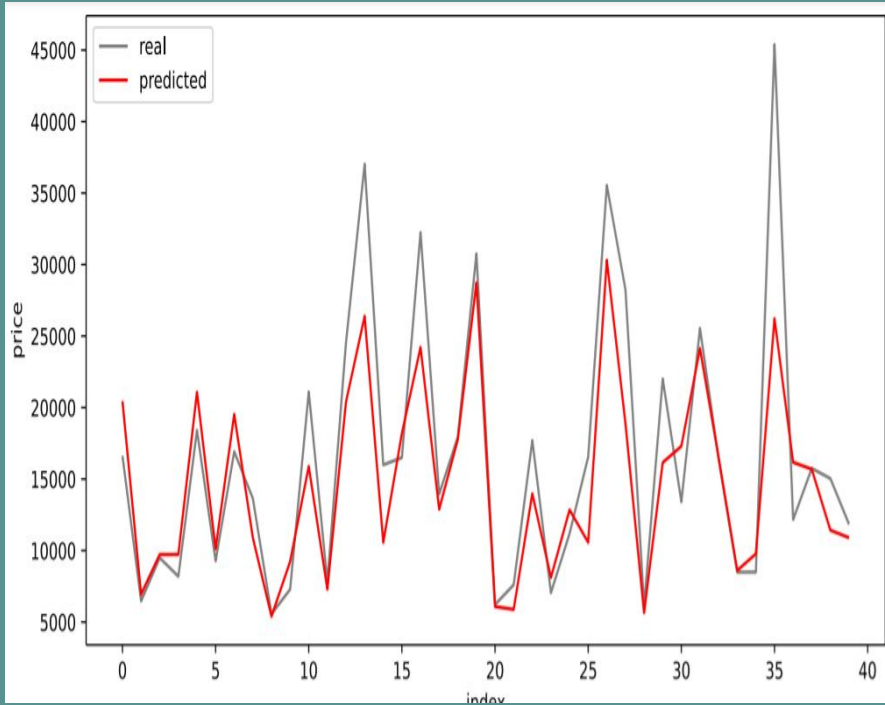
RMSE : 2114
R2 : 0,9325

LGBMR Regressor

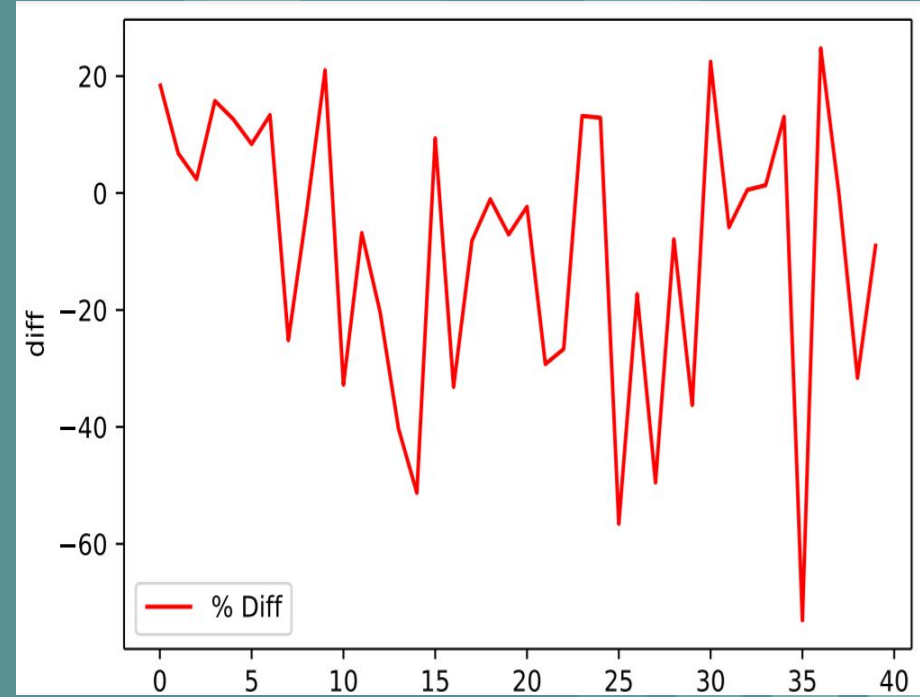
RMSE : 2439
R2 : 0,9102

Comparison

Real vs Predicted



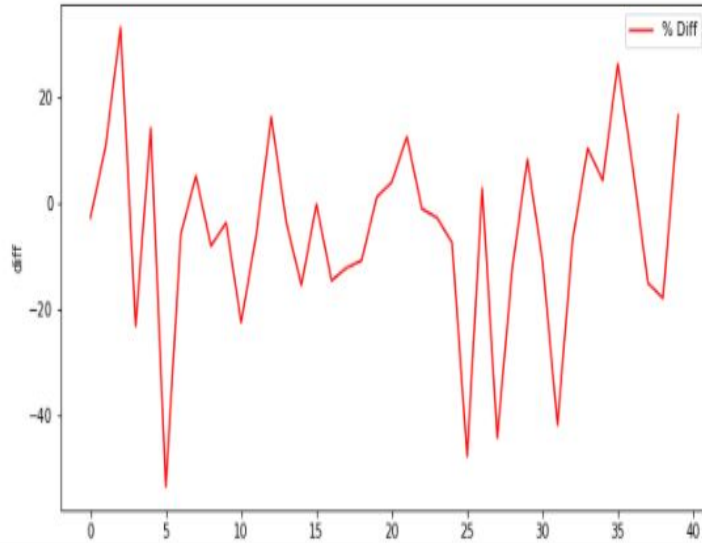
% Of Difference



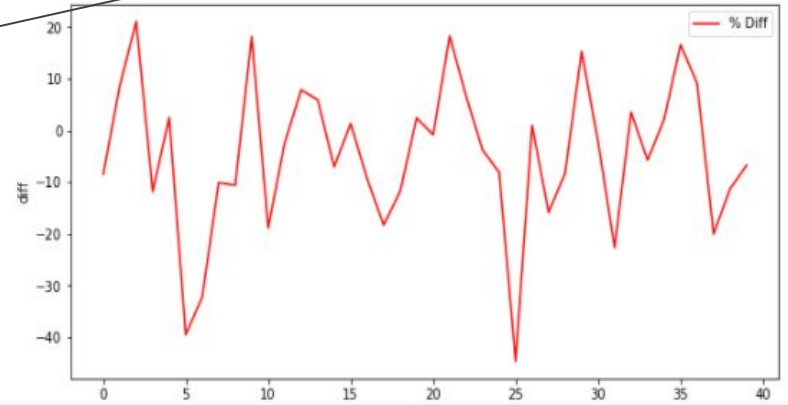
Comparison

Melhor modelo

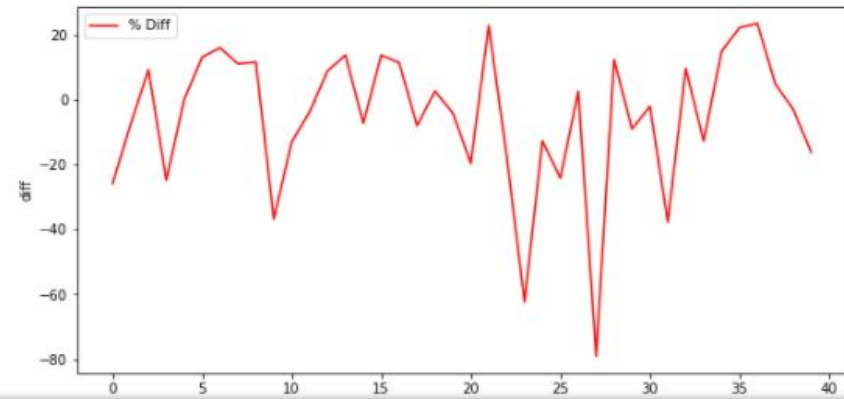
LGBMRegressor



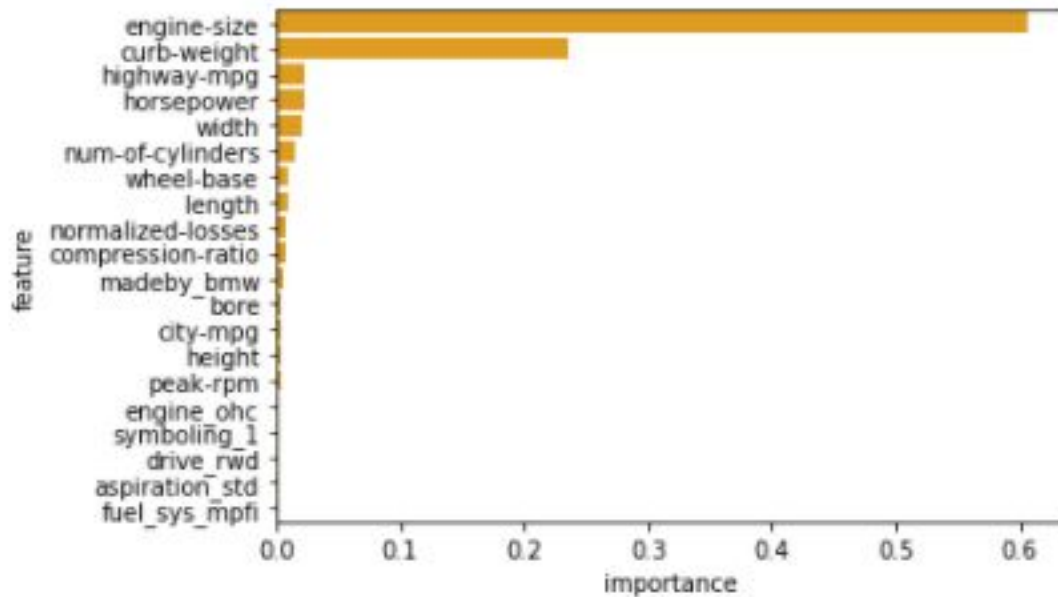
RandomForestRegressor



LinearRegression



Importance of Features



Model for APP

RMSE: 2009.744
R2: 0.939



Auto 1985 Predictor

We need some information to predict the car price

engine-size
50 350

curb-weight
1000 5000

highway-mpg
10 60

horsepower
30 300

width
50 100



Parameter	Selected Value	Range
engine-size	50	50 - 350
curb-weight	1000	1000 - 5000
highway-mpg	10	10 - 60
horsepower	30	30 - 300
width	50	50 - 100

The estimated price is \$5551.72

TRY HERE