**MACHINE LEARNING**

**Dengan Metode Liner Regression**

1. **Analisis data**
2. Import library
3. Numpy

Library python yang focus pada *scientific computing.* Numpy memiliki kemampuan untuk membentuk objek N-dimensional array, yang mirip dengan list pada python.

import numpy as np

1. Pandas

Library analisis data yang memiliki struktur data yang diperlukan untuk membersihkan data mentah ke dalam sebuah bentuk yang cocok untuk analisis (yaitu tabel)

import pandas as pd

1. Matplotlib

Digunakan untuk visualisasi dataset kedalam bentuk gambar

import matplotlib.pyplot as plt

1. Seaborn

Digunakan untuk visualisasi dataset kedalam bentuk gambar dan menghasilkan visualisasi yang lebih menarik

import seaborn as sns

1. LinearRegression

Analisis statistic yang digunakan untuk mengetahui pengaruh beberapa variable bebas (independent) terhadap variable terikat (dependent)

from sklearn.linear\_model import LinearRegression

1. train\_test\_split

Fungsi dalam pemilihan model Sklearn untuk membagi array data menjadi dua subset: untuk data pelatihan dan untuk data pengujian

from sklearn.model\_selection import train\_test\_split

1. Membaca datafile

Untuk membaca data yang sudah dijadikan dalam 1 folder

df = pd.read\_csv('cardata.csv')

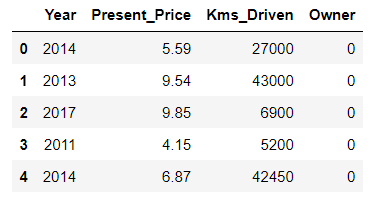
Untuk mengambil beberapa kolom saja

df = pd.read\_csv('cardata.csv', usecols=['Present\_Price','Year','Kms\_Driven','Owner'])

1. Menampilkan datafile

df.head()

Output :



1. Untuk mengetahui jumlah baris dan kolom datafile kita

df.shape

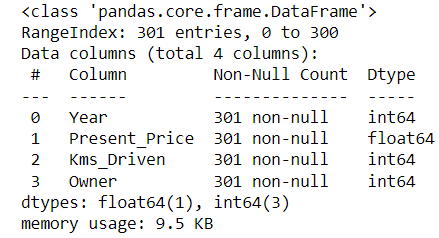
Output :



1. Untuk menampilkan info datafile kita meliputi baris, kolom, tipe data, memori yg digunakan, dll

df.info()

Output :



1. Untuk melihat statiska datafile kita meliputi rata-rata, minimum, maximum, dll

df.describe()

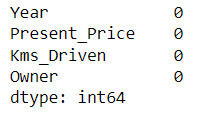
Output :



1. Untuk mengetahui missing value atau nilai yg kosong

df.isnull().sum()

Output :



1. Analisis untuk melihat distribusi dari masing-masing kolom dalam bentuk grafik dan box

f = plt.figure(figsize=(20,10))

f.add\_subplot(1,2,1)

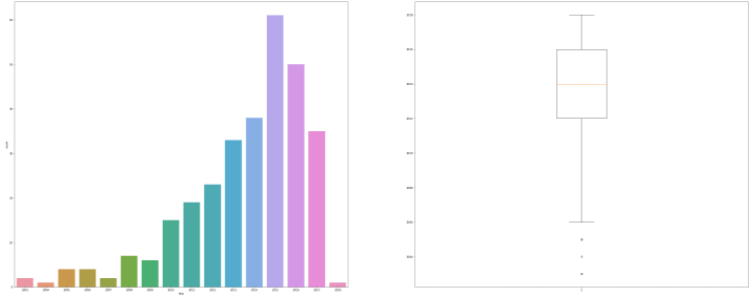
sns.countplot(df['Year'])

f.add\_subplot(1,2,2)

plt.boxplot(df['Year'])

plt.show()

Output :



1. Untuk melihat sencity/pemerataan distribusi dari harga

f = plt.figure(figsize=(12,4))

f.add\_subplot(1,2,1)

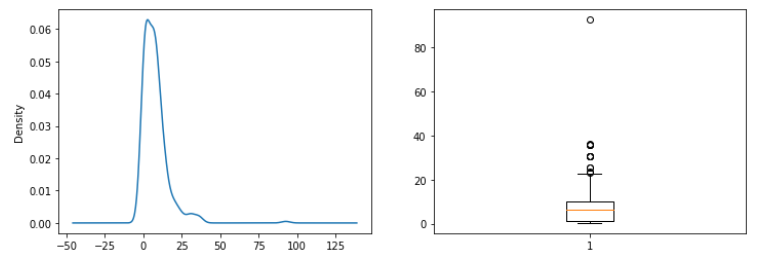
df['Present\_Price'].plot(kind='kde')

f.add\_subplot(1,2,2)

plt.boxplot(df['Present\_Price'])

plt.show()

Output :



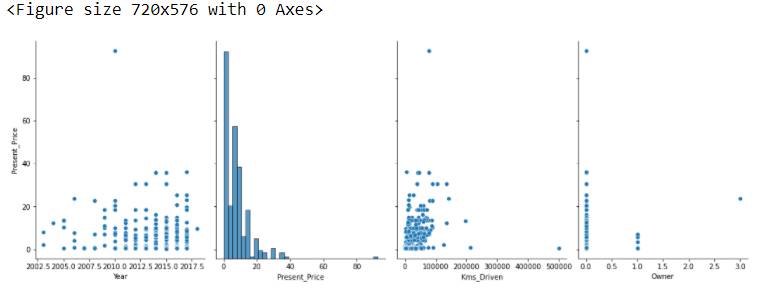
1. Untuk mengetahui hubungan antara independen dan dependen ariabel

f = plt.figure(figsize=(10,8))

sns.pairplot(data=df, x\_vars=['Year','Present\_Price','Kms\_Driven','Owner'], y\_vars=['Present\_Price'], size=5, aspect=0.75)

plt.show()

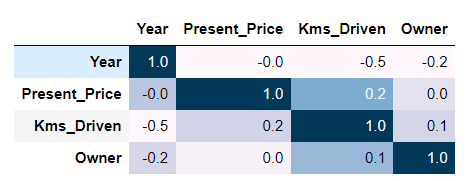
Output :



1. Untuk mengetahui nilai korelasi dari independent dan dependen variable

df.corr().style.background\_gradient().set\_precision(1)

Output :



1. **Modelling data**
2. Membagi variabel independen dan dependen menggunakan x dan y

x = df.drop(columns='Present\_Price')

y = df['Present\_Price']

1. Membuat data training dan testing

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.2, random\_state=0)

1. Untuk mengetahui jumlah baris dan kolom masing-masing variable

print(x\_train.shape)

Output :



1. Buat objek variabel linear regression

lr = LinearRegression()

1. Fit variabel x & y train

lr.fit(x\_train, y\_train)

Output :



1. Untuk print nilai koefisien

print(lr.coef\_)

print(lr.intercept\_)

Output :



1. Untuk menampilkan data dalam bentuk table

coef\_dict = {

'features':x.columns,

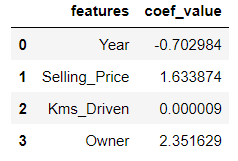
'coef\_value':lr.coef\_

}

coef = pd.DataFrame(coef\_dict, columns=['features','coef\_value'])

coef

Output :



1. Untuk memprediksi x dan y

y\_pred = lr.predict(x\_test)

1. Untuk mencari akurasi score dari model kita

lr.score(x\_test, y\_test)

Output :



Nilai akurasi score kita 0.8816045479359077 = 88,16% yang artinya bagus

1. Sudah bisa Melakukan prediksi sesuai permintaan

Contoh :

Ali ingin membeli mobil dengan spesifikasi

* Tahun 2017
* Selling price 6.0
* Kms Driven 10000
* Owner 0

Maka berapakah uang yang akan disiapkan ali untuk membeli mobil tersebut?

lr.predict([[2017,6.0,10000,0]])

Output :



Yang berarti Ali harus menyyiapkan sekitar 7.2821 dollar untuk membeli mobil dengan spesifikasi yang diingkannya