Decision Tree for Regression

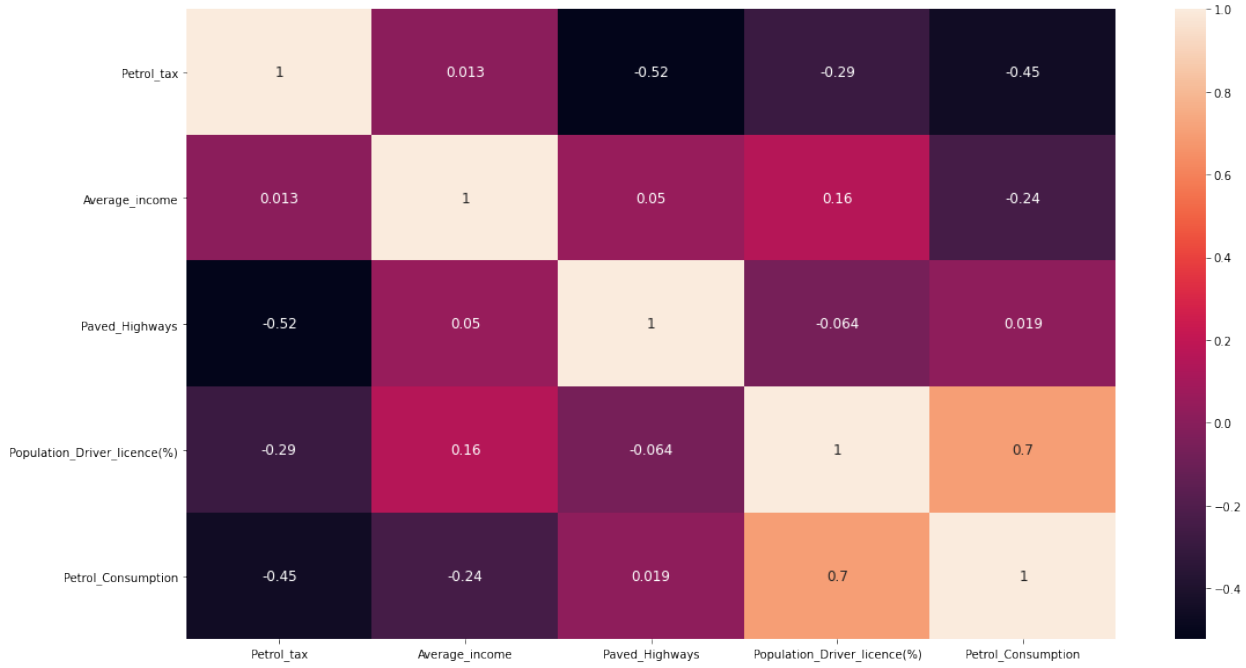
```
path='https://frenzy86.s3.eu-west-2.amazonaws.com/fav/tecno/
petrol_consumption.csv'
```

```
   Petrol_tax  Average_income  ...  Population_Driver_licence(%)
Petrol_Consumption
0         9.0            3571  ...                         0.525
541
1         9.0            4092  ...                         0.572
524
2         9.0            3865  ...                         0.580
561
3         7.5            4870  ...                         0.529
414
4         8.0            4399  ...                         0.544
410

[5 rows x 5 columns]
```

```
       Petrol_tax  ...  Petrol_Consumption
count   48.000000  ...           48.000000
mean     7.668333  ...          576.770833
std      0.950770  ...          111.885816
min      5.000000  ...          344.000000
25%      7.000000  ...          509.500000
50%      7.500000  ...          568.500000
75%      8.125000  ...          632.750000
max     10.000000  ...          968.000000

[8 rows x 5 columns]
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48 entries, 0 to 47
Data columns (total 5 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   Petrol_tax                     48 non-null     float64
 1   Average_income                 48 non-null     int64
 2   Paved_Highways                 48 non-null     int64
 3   Population_Driver_licence(%)   48 non-null     float64
```

```
 4   Petrol_Consumption            48 non-null     int64
dtypes: float64(2), int64(3)
memory usage: 2.0 KB
```

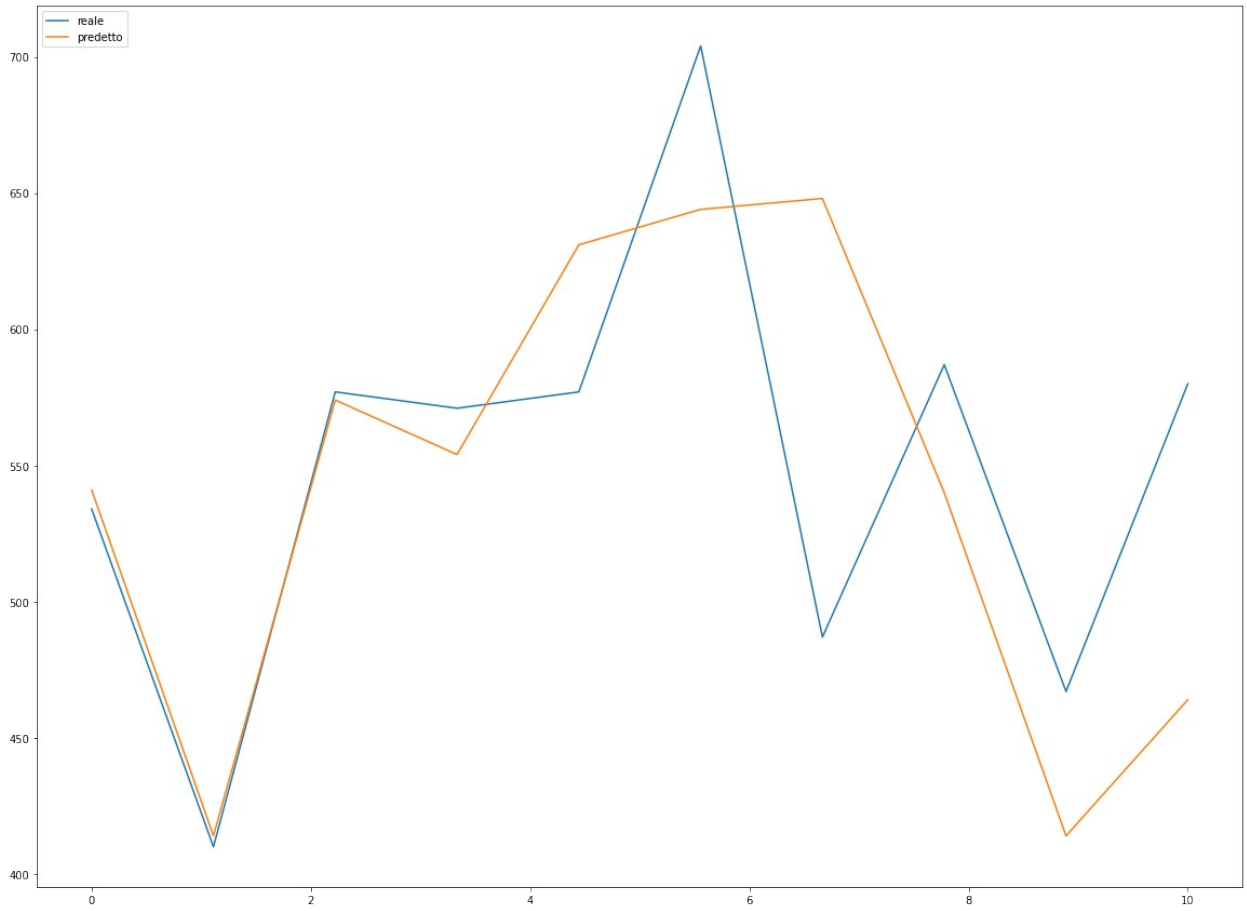


Regression

```
DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
                      max_features=None, max_leaf_nodes=None,
                      min_impurity_decrease=0.0,
min_impurity_split=None,
                      min_samples_leaf=1, min_samples_split=2,
                      min_weight_fraction_leaf=0.0,
presort='deprecated',
                      random_state=None, splitter='best')
```

```
    reale  predetto
29    534     541.0
4     410     414.0
26    577     574.0
```

```
30    571     554.0
32    577     631.0
37    704     644.0
34    487     648.0
40    587     540.0
7     467     414.0
10    580     464.0
```



Evaluating the Algorithm

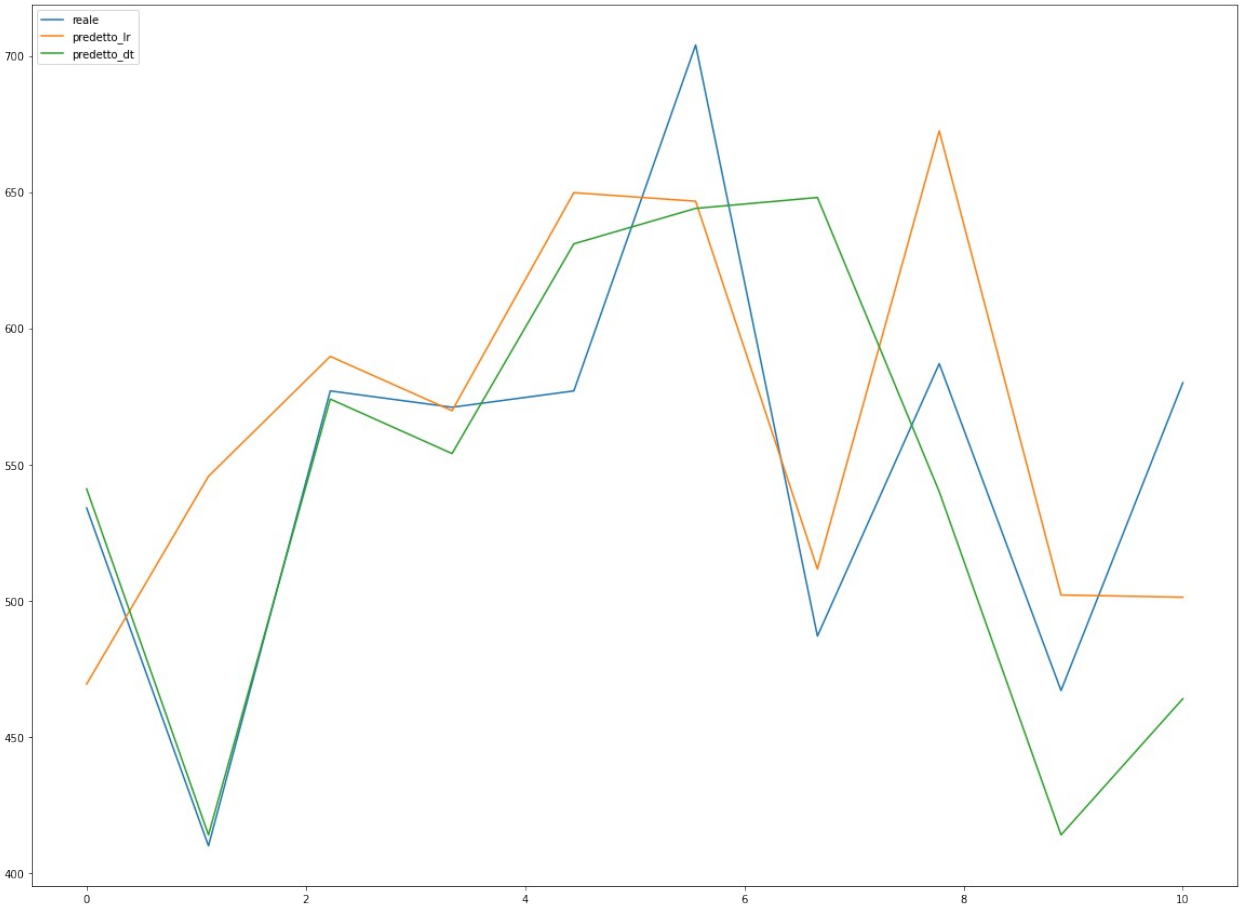
To evaluate performance of the regression algorithm, the commonly used metrics are mean absolute error, mean squared error, and root mean squared error.

```
Mean Absolute Error (MAE): 56.2
Mean Squared Error (MSE): 6298.2
Root Mean Squared Error (RMSE): 79.36119958770784
```

With Linear Regression

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
normalize=False)
```

	reale	predetto_lr
29	534	469.391989
4	410	545.645464
26	577	589.668394
30	571	569.730413
32	577	649.774809
37	704	646.631164
34	487	511.608148
40	587	672.475177
7	467	502.074782
10	580	501.270734



```
Mean Absolute Error (MAE): 56.8222474789647
Mean Squared Error (MSE): 4666.344787588362
Root Mean Squared Error (RMSE): 68.31064915215168
```