Project On: Fuel Consumption Ratings Regression

Created by:

Kuldeep Singh Shekhawat

Sachin Kanojiya

STEPS OF MODEL

- Importing Required Libraries
- Importing Dataset
- Data Preprocessing : Drop unwanted features
- Data Visualisation
- Data Preprocessing for Prediction
- Prediction : Model development

Importing Required Libraries

- Import pandas as pd
- import numpy as np
- import seaborn as sns
- import matplotlib.pyplot as plt
- %matplotlib inline
- sns.set_style('darkgrid')

IMPORTING DATASET

	Model Year	Make	Model	Vehicle Class	Engine Size(L)	Cylinders	Transmission	Fuel Type	Fuel Consumption (City (L/100 km)	Fuel Consumption(Hwy (L/100 km))	Fuel Consumption(Comb (L/100 km))	Fuel Consumption(Comb (mpg))	CO2 Emissions(g/km)	CO2 Rating	Smog Rating
0	2022	Acura	ILX	Compact	2.4	4	AM8	Z	NaN	7.0	8.6	33.0	200.0	6.0	3.0
1	2022	Acura	MDX SH-AWD	SUV: Small	3.5	6	AS10	Z	12.6	9.4	11.2	25.0	263.0	4.0	5.0
2	2022	Acura	RDX SH-AWD	SUV: Small	2.0	4	AS10	Z	11.0	8.6	9.9	29.0	232.0	5.0	6.0
3	2022	Acura	RDX SH-AWD A-SPEC	SUV: Small	2.0	4	AS10	Z	11.3	9.1	10.3	27.0	242.0	5.0	6.0
4	2022	Acura	TLX SH-AWD	Compact	2.0	4	AS10	Z	11.2	8.0	9.8	29.0	230.0	5.0	7.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
940	2022	Volvo	XC40 T5 AWD	SUV: Small	2.0	4	AS8	Z	10.7	7.7	9.4	30.0	219.0	5.0	5.0
941	2022	Volvo	XC60 B5 AWD	SUV: Small	2.0	4	AS8	Z	10.5	8.1	9.4	30.0	219.0	5.0	5.0
942	2022	Volvo	XC60 B6 AWD	SUV: Small	2.0	4	AS8	Z	11.0	8.7	9.9	29.0	232.0	5.0	NaN
943	2022	Volvo	XC90 T5 AWD	SUV: Standard	2.0	4	AS8	Z	11.5	8.4	10.1	28.0	236.0	5.0	5.0
944	2022	Volvo	XC90 T6 AWD	SUV: Standard	2.0	4	AS8	Z	NaN	8.9	10.8	26.0	252.0	5.0	7.0

945 rows × 15 columns

DATA INFO :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 945 entries, 0 to 944
Data columns (total 12 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Vehicle Class                         945 non-null    object
 1   Engine Size(L)                        945 non-null    float64
 2   Cylinders                             945 non-null    int64
 3   Transmission                          945 non-null    object
 4   Fuel Type                             944 non-null    object
 5   Fuel Consumption (City (L/100 km)     931 non-null    float64
 6   Fuel Consumption(Hwy (L/100 km))      930 non-null    float64
 7   Fuel Consumption(Comb (L/100 km))     945 non-null    float64
 8   Fuel Consumption(Comb (mpg))          931 non-null    float64
 9   CO2 Emissions(g/km)                   934 non-null    float64
 10  CO2 Rating                            932 non-null    float64
 11  Smog Rating                           927 non-null    float64
dtypes: float64(8), int64(1), object(3)
memory usage: 88.7+ KB
```

```
Vehicle Class                              0
Engine Size(L)                             0
Cylinders                                  0
Transmission                               0
Fuel Type                                  1
Fuel Consumption (City (L/100 km)         14
Fuel Consumption(Hwy (L/100 km))          15
Fuel Consumption(Comb (L/100 km))          0
Fuel Consumption(Comb (mpg))              14
CO2 Emissions(g/km)                       11
CO2 Rating                                13
Smog Rating                               18
dtype: int64
```

Finding Null values

Fill Null Values

```
Vehicle Class                              0
Engine Size(L)                             0
Cylinders                                  0
Transmission                               0
Fuel Type                                  0
Fuel Consumption (City (L/100 km)          0
Fuel Consumption(Hwy (L/100 km))           0
Fuel Consumption(Comb (L/100 km))          0
Fuel Consumption(Comb (mpg))               0
CO2 Emissions(g/km)                        0
CO2 Rating                                 0
Smog Rating                                0
dtype: int64
```

DATA PROCESSING : DROP UNWANTED FEATURES

Data preprocessing is the process of transforming raw data into anunderstandable format. It also makes the datasets more complete and efficientto perform data analysis

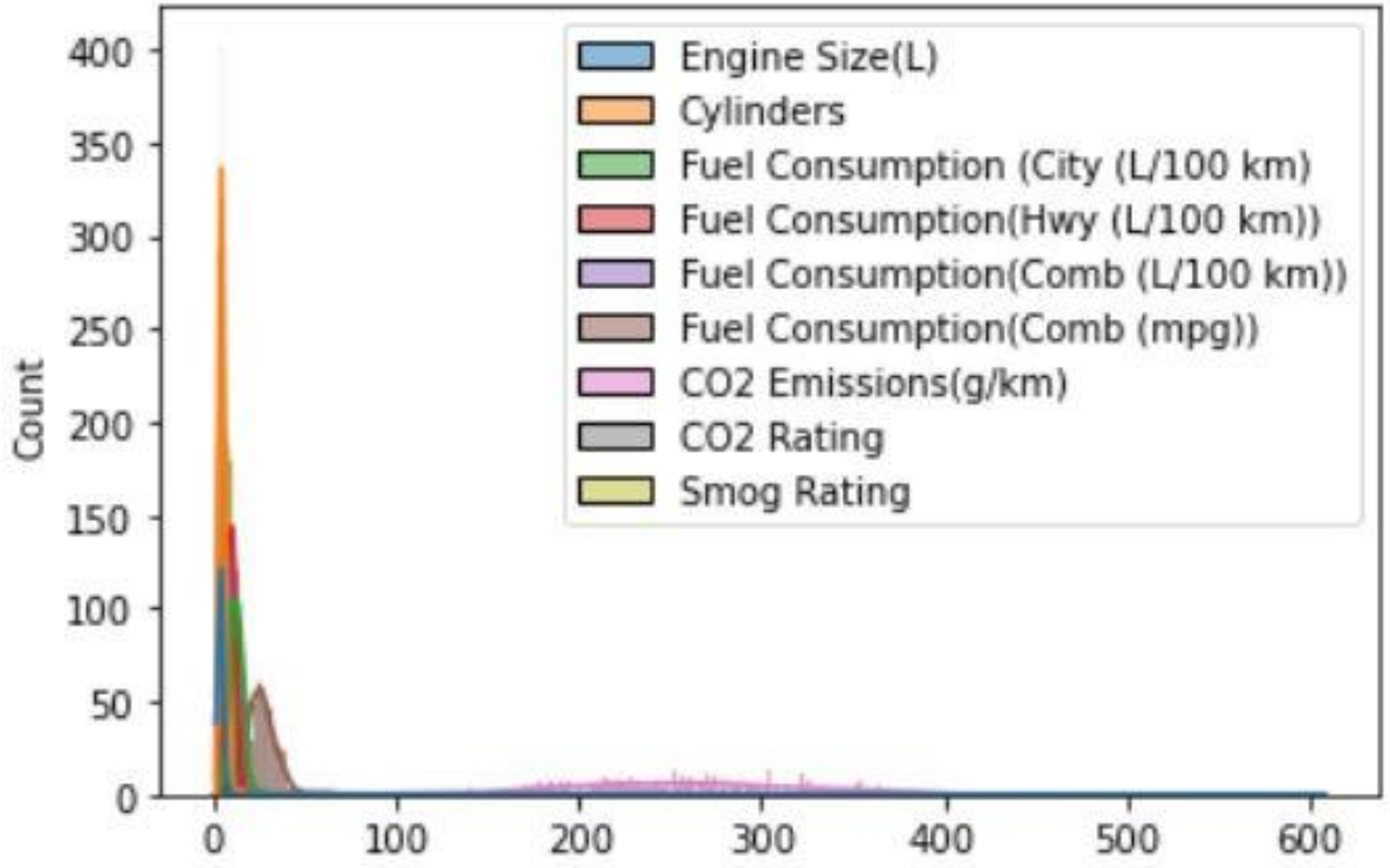
	Vehicle Class	Engine Size(L)	Cylinders	Transmission	Fuel Type	CO2 Emissions(g/km)	CO2 Rating	Smog Rating
0	Compact	2.4	4	AM8	Z	200.0	6.0	3.0
1	SUV: Small	3.5	6	AS10	Z	263.0	4.0	5.0
2	SUV: Small	2.0	4	AS10	Z	232.0	5.0	6.0
3	SUV: Small	2.0	4	AS10	Z	242.0	5.0	6.0
4	Compact	2.0	4	AS10	Z	230.0	5.0	7.0
...	...	...	...	...	...	...	...	...
940	SUV: Small	2.0	4	AS8	Z	219.0	5.0	5.0
941	SUV: Small	2.0	4	AS8	Z	219.0	5.0	5.0
942	SUV: Small	2.0	4	AS8	Z	232.0	5.0	5.0
943	SUV: Standard	2.0	4	AS8	Z	236.0	5.0	5.0
944	SUV: Standard	2.0	4	AS8	Z	252.0	5.0	7.0

945 rows × 8 columns

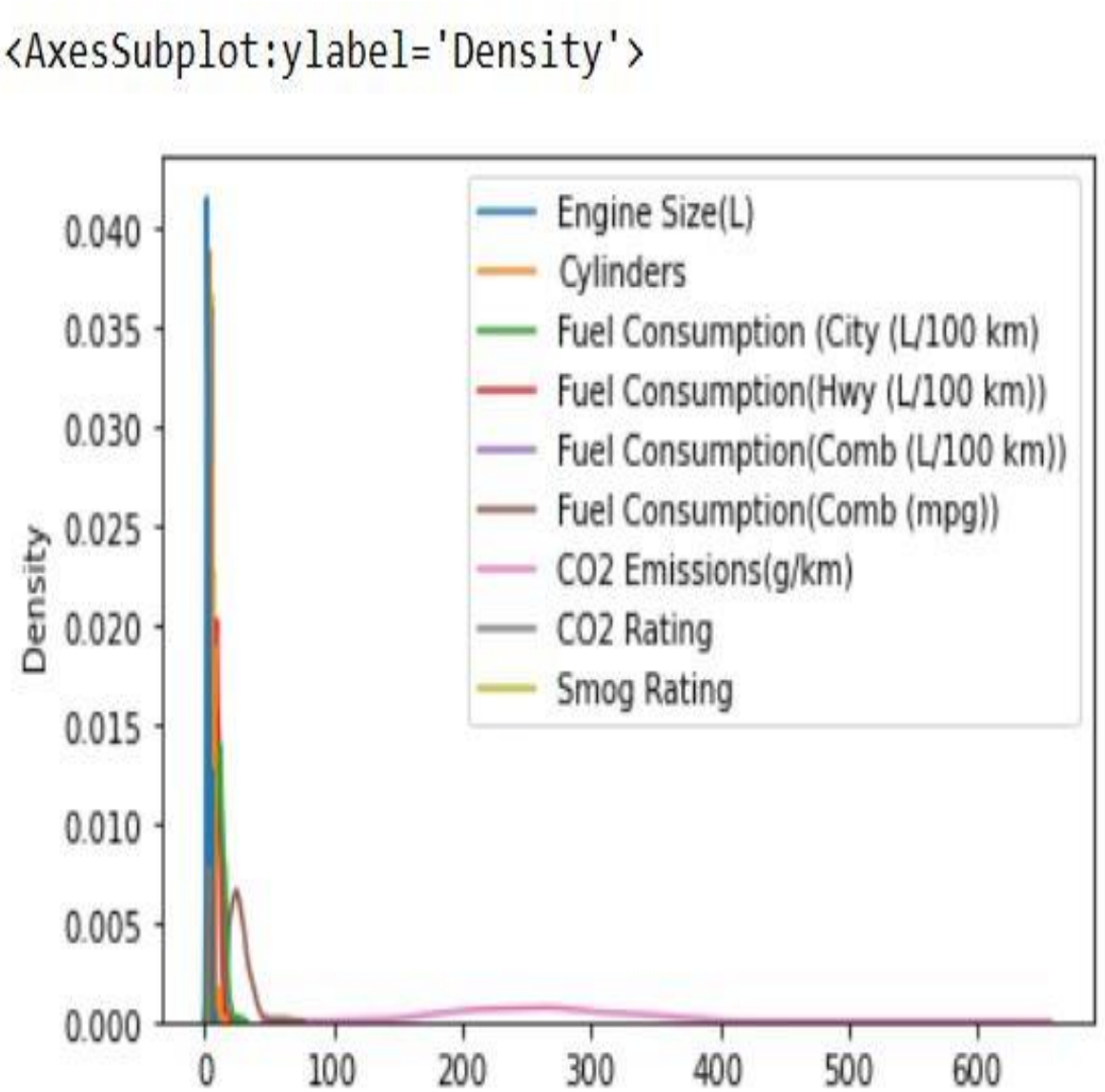
Data Visualisation

HISTPLOT

```
<AxesSubplot:ylabel='Count'>
```

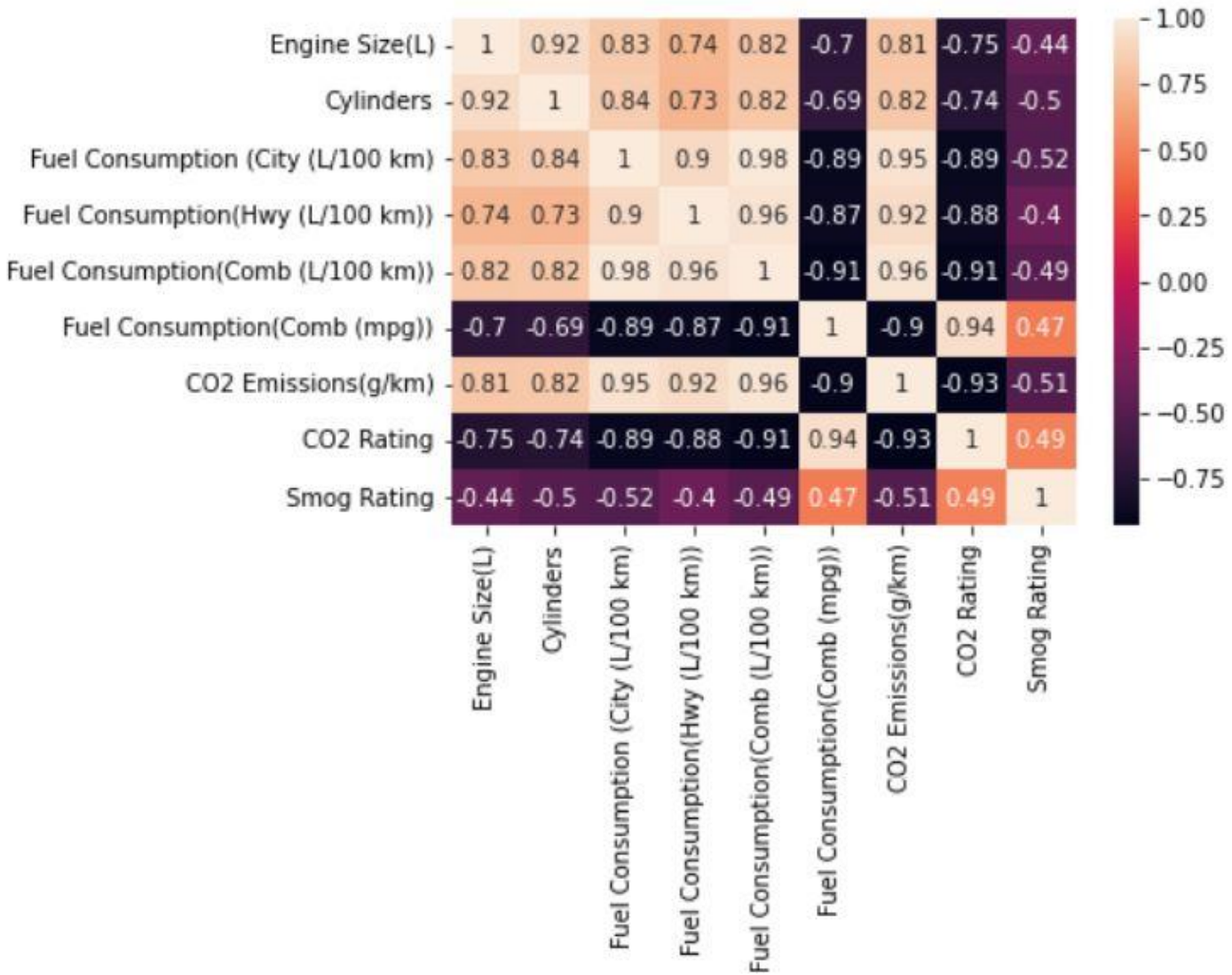


Kdeplot


<AxesSubplot:ylabel='Density'>
Density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
100
200
300
400
500
600
Engine Size(L)
Cylinders
Fuel Consumption (City (L/100 km)
Fuel Consumption(Hwy (L/100 km))
Fuel Consumption(Comb (L/100 km))
Fuel Consumption(Comb (mpg))
CO2 Emissions(g/km)
CO2 Rating
Smog Rating

heatmap

```
<AxesSubplot:>
```



Data Preprocessing for Prediction

Before using ML model the data is processed again in 3 steps

1.Encoding categorical features
2.Splitting test & train data

3.Feature Scaling

Encoding

- .Label Encoding refers to converting the labels into a numeric form so as to convert them into the machine readable form.We use it on our training data, and using “fit” it will figure out the unique values and assign a value to it, returns the encoded labels.
- Test-Train Split The test-train split is used to estimate the performance of machine learning algorithms using training data and validate with test data
- Feature scaling is a method used to normalize the range of independent variables or features of data and make them closer to each other.

- Logistic Regression Logistic Regression is a supervised learning algorithm that investigates the relationship b/w a dependent and independent variable and produces results in abinary format which is used to predict the out come of a categorical variable (Failue or not).

- SVM Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. The algorithm creates a line or ahyperplane which separates the data into classes.

Linear Regression Model

0	0.822622
1	-0.036975
2	0.093616
3	0.585306
4	0.804153
...	...
279	0.191900
280	0.481637
281	0.194040
282	-0.127257
283	-1.120876

284 rows × 1 columns

Thank You