

NAME OF THE PROJECT

Used Car Price Prediction

Submitted by:

Bishwajit Bhattacharya

**ACKNOWLEDGMENT**

I web scrap used car data from cars24.com I took data from 25 cities.

Total in my data set has 515 used car data which has almost 215 different

Model of cars. I added emi/month also , the interest things is it is different in respect of cities.

After getting data I made this all on 1 excel .

**INTRODUCTION**

In pandemic situation all the small business field is affected even car industries.

Statistics report says car sell counts in all over the world is down , but interestingly used car business is growing. But we have to know actual used cars price if we want to grow a business , for metropolitan cities this data we can get easily but for small cities it is very tough to get.

If you saw my data I got the used cars details not only from big city like Delhi, Mumbai, Bangalore, I got the data from city like Mysore, Ghaziabad, Jaipur, Rajkot etc. Although its fact that we can not get most luxury model in this cities but it is standard and easy emi available so it can easy to process.

First stapes of getting data is to webscrap for website.

I used both selenium and beautifulsup for this project.

Selenium for first stage to get correct website,

For the car name the code is

cars\_list = driver.find\_elements\_by\_xpath("//div[@class='\_1jpRU']")

cars = []

for i in cars\_list:

a = i.get\_attribute('innerText')

print(a)

cars.append(a)

print(cars)

For Price

price\_list = driver.find\_elements\_by\_xpath("//div[@class='\_7udZZ']")

price = []

for i in price\_list:

a = i.get\_attribute('innerText')

print(a)

price.append(a)

print(price)

For Emi

emi\_list = driver.find\_elements\_by\_xpath("//div[@class='\_2HFRN']")

emi = []

for i in emi\_list:

a = i.get\_attribute('innerText')

print(a)

emi.append(a)

print(emi)

For Beutifulsup the codes are

scraped\_cars = soup.find\_all('h2' , class\_='\_3FpCg')

scraped\_cars

cars = []

for car in scraped\_cars:

car = car.get\_text().replace('\n',"")

car = car.strip(" ")

cars.append(car)

cars

scraped\_price = soup.find\_all('div' , class\_='\_7udZZ')

scraped\_price

prices = []

for price in scraped\_price:

price = price.get\_text().replace('\n',"")

price = price.strip(" ")

prices.append(price)

prices

scraped\_emi = soup.find\_all('div' , class\_='\_2HFRN')

scraped\_emi

emies = []

for emi in scraped\_emi:

emi = emi.get\_text().replace('\n',"")

emi = emi.strip(" ")

emies.append(emi)

emies.

After getting al the data I took all data on 1 single excel “**Carprediction.xls**”

Now the first step of data cleaning is to crate data set.

df = pd.DataFrame(pd.read\_excel("Carprediction.xlsx"))

df

we have total 514 rows and 5 columns target columns is price(rs)

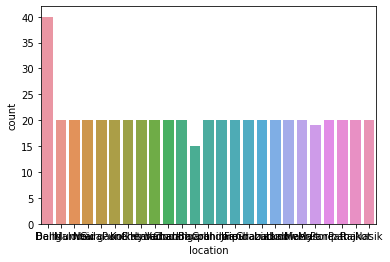
Then we need to check is there any null value.

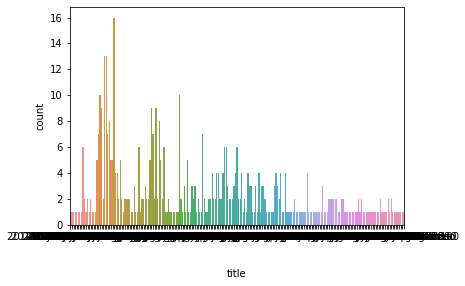
df.isnull().sum()

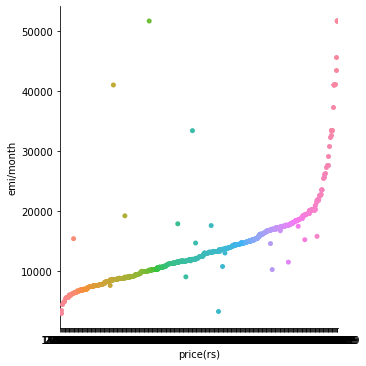
then we need to check what is the data type

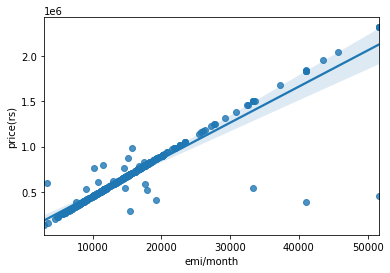
df.dtypes

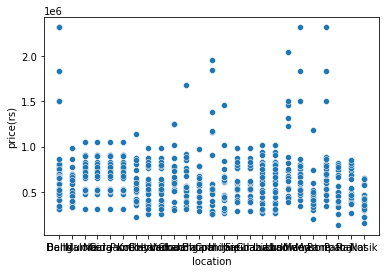
observation is target columns price(rs) is int so we will process regression techniques but first data visualization. I plot different types of graph like count plot, catplot, regplot, scatterplot and lmpot.

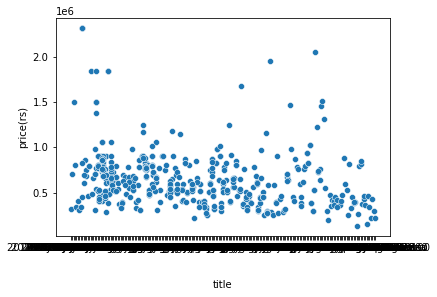


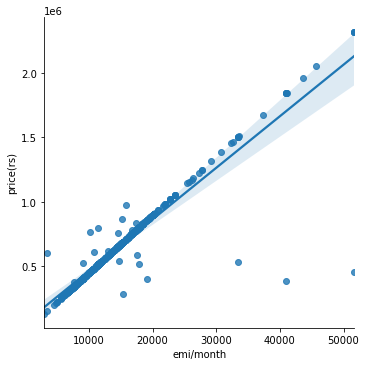












Now we change the object to float

for i in df.columns:

if df[i].dtypes=="object":

df[i]=enc.fit\_transform(df[i].values.reshape(-1,1))

Heatmap

<AxesSubplot:>



So now we will process regression Technique.

I used linear, lasso and ridge regression techniques,

Then I done the ensamble technique bye random forest techniques.

After this I got

R2 Score: 99.9287627889498

Cross Val Score: 98.34741615127285

Mean absolute error:: 1.1552049639454532e-13

Mean squared error:: 2.3577149012033123e-26

Root mean square:: 1.535485233144009e-13

Conclusion:-

I save the model in carpresiction.pkl

Got the result of

0.999287627889498

So bye using this model we can get almost 99% accurate price of used car.

The full code is in

https://github.com/bishwa2017/Car-Prediction/blob/main/Carpredictionproject.ipynb