CAPSTONE PROJECT WORK REPORT

Phase II

CAR PRICE PREDICTION

BALAJI S

Supervisor: **Mrs.P.Jayapriya**, MCA.,M.E.,(Ph.D.)
, Dept. of CS with DA

Department of Computer Science with Data Analytics

KPR College of Arts Science and Research
(Affiliated to Bharathiar University, Coimbatore)
Avinashi Road, Arasur, Coimbatore – 641 407

NOVEMBER 2022

CAPSTONE PROJECT WORK REPORT

Phase II

CAR PRICE PREDICTION

Bonafide Work Done by

BALAJI S

REG. NO. 2028B0006

Dissertation submitted in partial fulfillment of the requirements for the award of Bharathiar University, Coimbatore-46.

Signature of the Guide **Signature of the HOD**

[Mrs.P.Jayapriya]

Submitted for the Viva-Voce Examination held on _____

Internal Examiner **External Examiner**

TABLE OF CONTENTS **PAGE NO.**

ACKNOWLEDGEMENT

In the accomplishment of completion of my Capstone Project Work Phase- II on CAR PRICE PREDICTION USING MACHINE LEARING I would like to convey my special gratitude to**Dr. S. Balusamy**, **Principal of KPR College of Arts Science and Research** and **Mrs.P.Jayapriya, Associate Professor ,Department of Computer Science with Data Analytics.**Your valuable guidance and suggestions helped me in phase - II of the completion of this project. I will always be thankful to you in this regard. I am ensuring that this project was and not copied.

Student Signature

Place :

Date:

KPR COLLEGE OF ARTS SCIENCE AND RESEARCH
(Affiliated to Bharathiar University, Coimbatore)
Avinashi Road,Arasur,Coimbatore-641407

ABOUT THE COLLEGE

KPR College of Arts Science and Research is the latest addition to the KPR fleet. The College is located in a picturesque campus of about 11. Acres. The College is run by KPR charities under the leadership of our Chairman Dr. K.P. Ramasamy. The KPR Group is one of the largest industrial conglomerate in the country with interest in Textiles, Sugar, Wind Turbines, Automobiles and Education. The College was established in the year 2019 with a vision of providing top class education and life skills to students and thereby serve the nation and beyond. KPRCAS today offers 12 UG programmes in Management, Commerce and Computer Science streams. The Students of KPRCAS undergo intense training not only in the syllabus and curriculum of the affiliating University but are also trained in various areas. So that they emerge as industry ready graduates to meet the varying demands of the competing industries. Character building and Leadership qualities are inculcated into the students to make them responsible citizens focusing on the development of society and nation. A plethora of Clubs and Events encouraged the students to take part in sports and other cultural activities. KPRCAS offers three years undergraduate courses, which are exclusively for Business, Commerce and Computer Science Stream. The students are equipped with skills and knowledge needed to take up various leadership positions and to develop the society. Beyond Book Teaching help them to be professionals. KPRCAS emphasis on making the students academically brilliant, and also prepare them for the real corporate world. The learning curve begins here for the students of KPRCAS.

ABOUT THE DEPARTMENT

Bachelor of Computer Science with Data Analytics (B.Sc. (CS with DA)) was established in the year 2020. Data Analytics helps to raise the quality of data in the entire business system. The goal of data analytics is to construct the means for extracting businessfocused insights from data This requires an understanding of how value and information flows in a business, and the ability to use that understanding to identify business opportunities. The primary aim of a data analyst is to increase efficiency and improve performance by discovering patterns in data. Data analysts exist at the intersection of information technology, statistics and business. They combine these fields in order to help businesses and organizations succeed. The students get exposed to Big Data, Business Intelligence, Data Mining, Data Visualization, Advanced Excel, Predictive Analytics and R Programming.

SYNOPSIS

A car price prediction has been a highinterest research area, as it requires noticeable effort and knowledge of the field expert. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, we applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal autopijaca.ba using web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction model was integrated into Java application. Furthermore, the model was evaluated using test data and the accuracy of 87.38% was obtained.

CHAPTER 1

1.Introduction

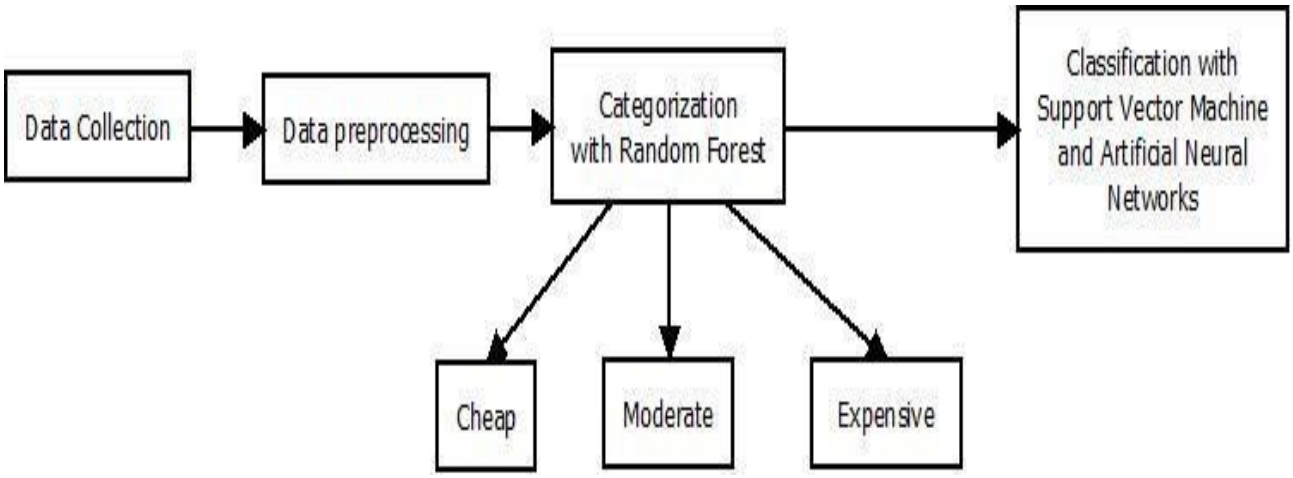
Car price prediction is somehow interesting and popular problem. As per information that was gotten from the Agency for Statistics of BiH, 921.456 vehicles were registered in 2014 from which 84% of them are cars for personal usage [1]. This number is increased by 2.7% since 2013 and it is likely that this trend will continue, and the number of cars will The article is published with Open Access at www.temjournal.com increase in future. This adds additional significance to the problem of the car price prediction. Accurate car price prediction involves expert knowledge, because price usually depends on many distinctive features and factors. Typically, most significant ones are brand and model, age, horsepower and mileage. The fuel type used in the car as well as fuel consumption per mile highly affect price of a car due to a frequent changes in the price of a fuel. Different features like exterior color, door number, type of transmission, dimensions, safety, air condition, interior, whether it has navigation or not will also influence the car price. In this paper, we applied different methods and techniques in order to achieve higher precision of the used car price prediction.

1.1.Related Work

Predicting price of a used cars has been studied extensively in various researches. Listian discussed, in her paper written for Master thesis [2], that regression model that was built using Support Vector Machines (SVM) can predict the price of a car that has been leased with better precision than multivariate regression or some simple multiple regression. This is on the grounds that Support Vector Machine (SVM) is better in dealing with datasets with more dimensions and it is less prone to overfitting and underfitting. The weakness of this research is that a change of simple regression with more advanced SVM regression was not shown in basic indicators like mean, variance or standard deviation.

1.2 Materials and Methods

Approach for car price prediction proposed in this paper is composed of several steps, shown in



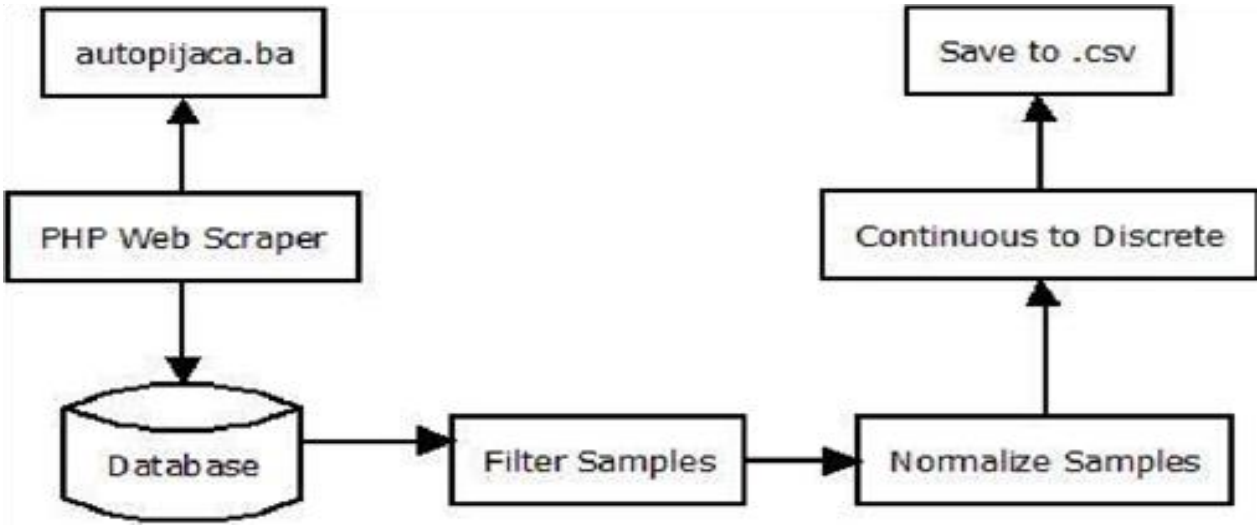
.

Data is collected from a local web portal for selling and buying cars autopijaca.ba [9], during winter season, as time interval itself has high impact on the price of the cars in Bosnia and Herzegovina. The following attributes were captured for each car: brand, model, car condition, fuel, year of manufacturing, power in kilowatts, transmission type, millage, color, city, state, number of doors, four wheel drive (yes/no), damaged (yes/no), navigation (yes/no), leather seats (yes/no), alarm (yes/no), aluminum rims (yes/no), digital air condition (yes/no), parking sensors (yes/no), xenon lights (yes/no), remote unlock (yes/no), electric rear mirrors (yes/no), seat heat (yes/no), panorama roof (yes/no), cruise control (yes/no), abs (yes/no), esp (yes/no), asr (yes/no) and price expressed in BAM (Bosnian Mark).

Since manual data collection is time consuming task, especially when there are numerous records to process, a “web scraper” as a part of this research is created to get this job done automatically and reduce the time for data gathering. Web scraping is well known technique to extract information from websites and save data into local file or database. Manual data

extraction is time consuming and therefore web scrapers are used to do this job in a fraction of time. Web scrapers are programed for specific websiteand can mimic regular users from website's point of

.



Data gathering and transformation workflow diagram

brand	model	fuel	power in kilowatts	year of man	miles	leather	cruise control	price
volkswagen	golf2	Diesel	45-55	l7	l4	no	no	0-1500
volkswagen	golf2	Gasoline	0-45	l7	l4	no	no	0-1500
ford	escort	Gasoline	45-55	l7	l1	no	no	0-1500
ford	fiesta	Gasoline	55-65	l4	l2	no	no	0-1500

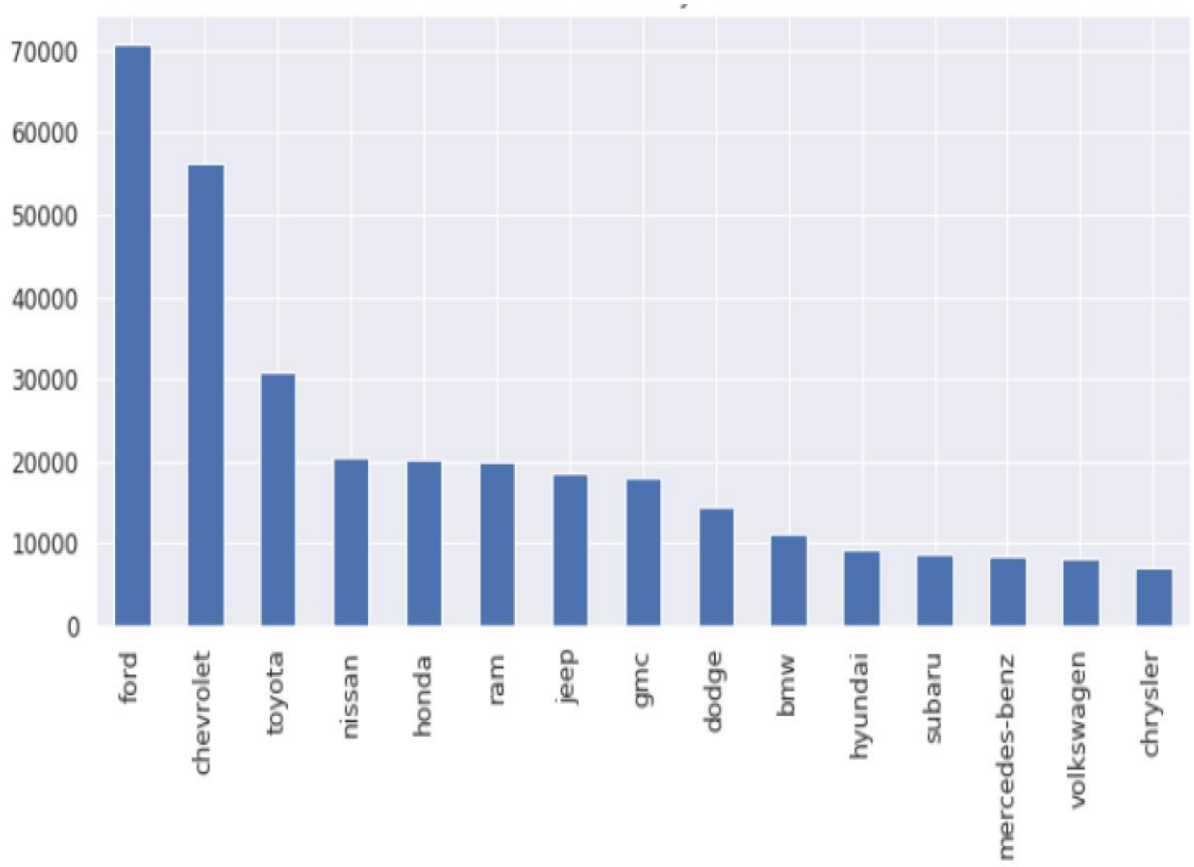
mercedes-benz	190	Gasoline	45-55	l7	l4	no	no	0-1500
volkswagen	jetta	Diesel	0-45	l7	l5	no	no	0-1500
ford	focus	Gasoline	55-65	l6	l4	no	no	0-1500
fiat	punto	Diesel	65-75	l5	l4	no	no	0-1500
volkswagen	golf2	Gasoline	65-75	l7	l4	no	no	0-1500

The color of the cars was normalized into fixed set of 15 different colors. Continuous attributes such as “millage”, “year of manufacturing”, “power in kilowatts” and “price” are converted into categorical values using predefined cluster intervals. The millage is converted into five Distinct categories, the year of manufacturing has been converted into seven categories and the power in kilowatts is converted into eleven categories. The price attribute has been categorized into 15 distinct categories based on price range. These categories are shown in Table 2 and similar principle was applied to other attributes. This data transformation process converted regression prediction machine learning problem into classification problem

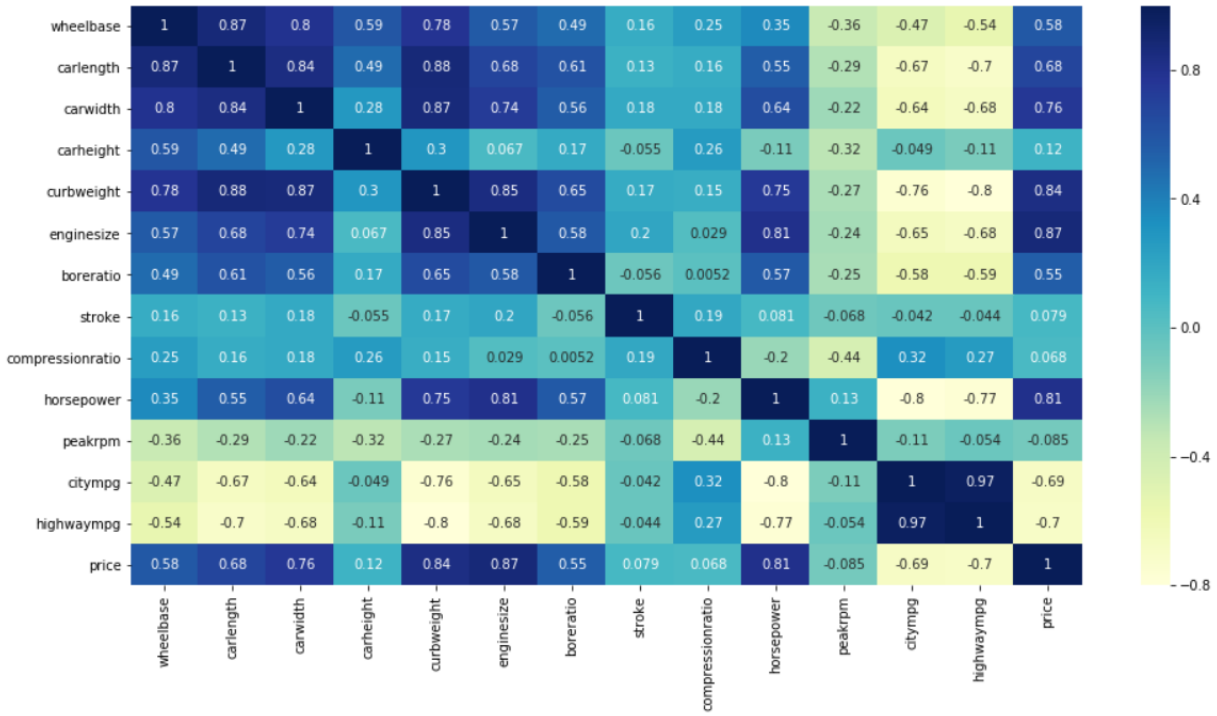
CHAPTER 2

2. SYSTEM STUDY

2.1 EXISTING SYSTEM



The column"**Price**" is the target variable and rest of the columns are independent variables.

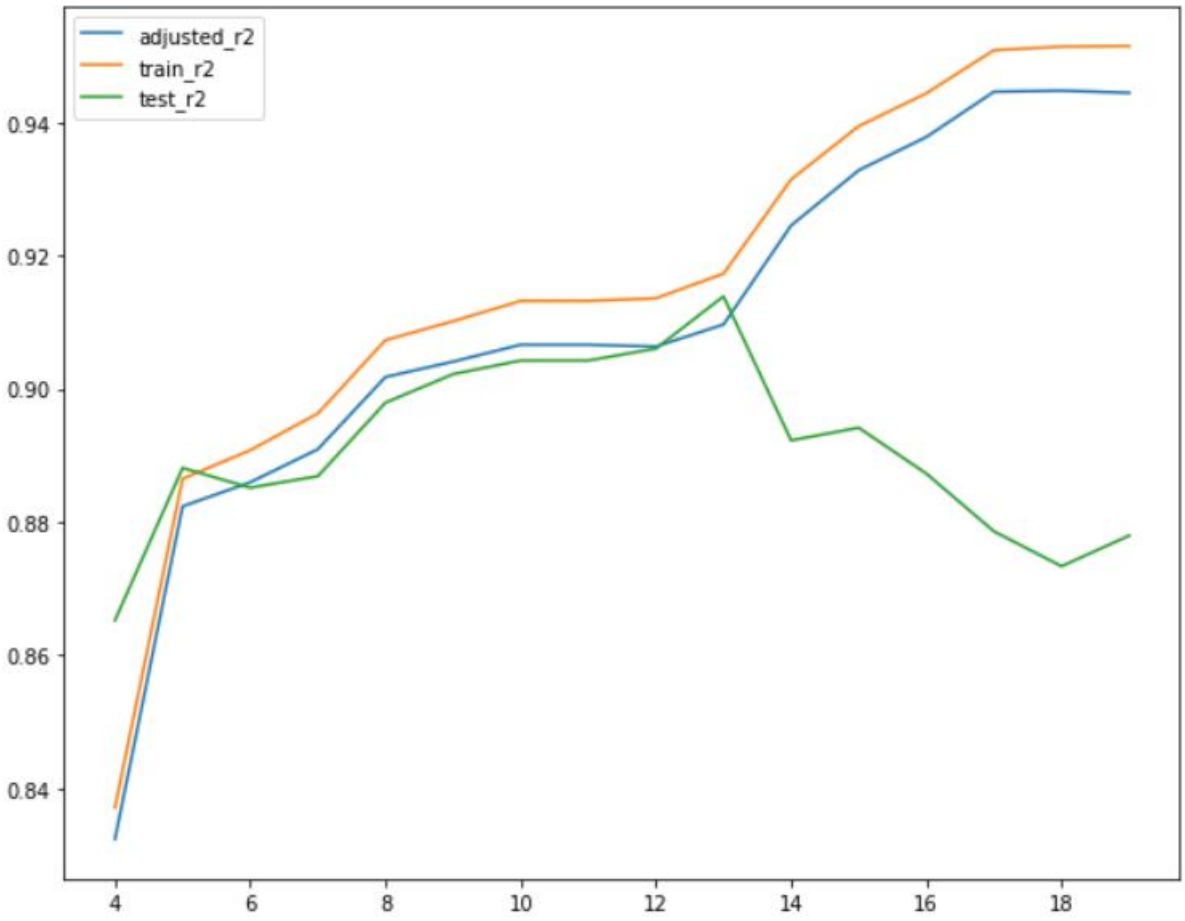


The heatmap shows some useful insights:

- any independent variables are highly correlated (look at the top-left part of matrix): wheelbase, carlength, curbweight, enginesize etc. are all measures of ‘size/weight’, and are positively correlated

Thus, while building the model, we’ll have to pay attention to multicollinearity (especially linear models, such as linear and logistic regression, suffer more from multicollinearity).

Data Preparation:



umber of Features vs R-squared

2.2 Drawbacks:

The prices of new cars in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But due to the increased price of new cars and the incapability of customers to buy new cars due to the lack of funds, used cars sales are on a global increase (Pal, Arora and Palakurthy, 2018). There is a need for a used car price prediction system to

effectively determine the worthiness of the car using a variety of features. Even though there are websites that offers this service, their prediction method may not be the best. Besides, different models and systems may contribute on predicting power for a used car's actual market value. It is important to know their actual market value while both buying and selling.

2.3 PROPOSED SYSTEM

One of the main areas of research in machine learning is the prediction of the price of cars. It is based on **finance** and the **marketing** domain. It is a major research topic in machine learning because the price of a car depends on many factors. Some of the factors that contribute a lot to the price of a car are:

1. Brand
2. Model
3. Horsepower
4. Mileage
5. Safety Features
6. GPS and many more.

Car Price Prediction Model using Python

The dataset I'm using here to train a car price prediction model was downloaded from Kaggle. It contains data about all the main features that contribute to the price of a car. So let's start this task by importing the necessary Python libraries and the

Dataset

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor

data = pd.read_csv("CarPrice.csv")
data.head()
```

car_ID	symboling	CarName	...	citympg	highwaympg	price
0 1	3	alfa-romero giulia	...	21	27	13495.0
1 2	3	alfa-romero stelvio	...	21	27	16500.0
2 3	1	alfa-romero Quadrifoglio	...	19	26	16500.0
3 4	2	audi 100 ls	...	24	30	13950.0
4 5	2	audi 100ls	...	18	22	17450.0
[5 rows x 26 columns]						

1

data.isnull().sum()

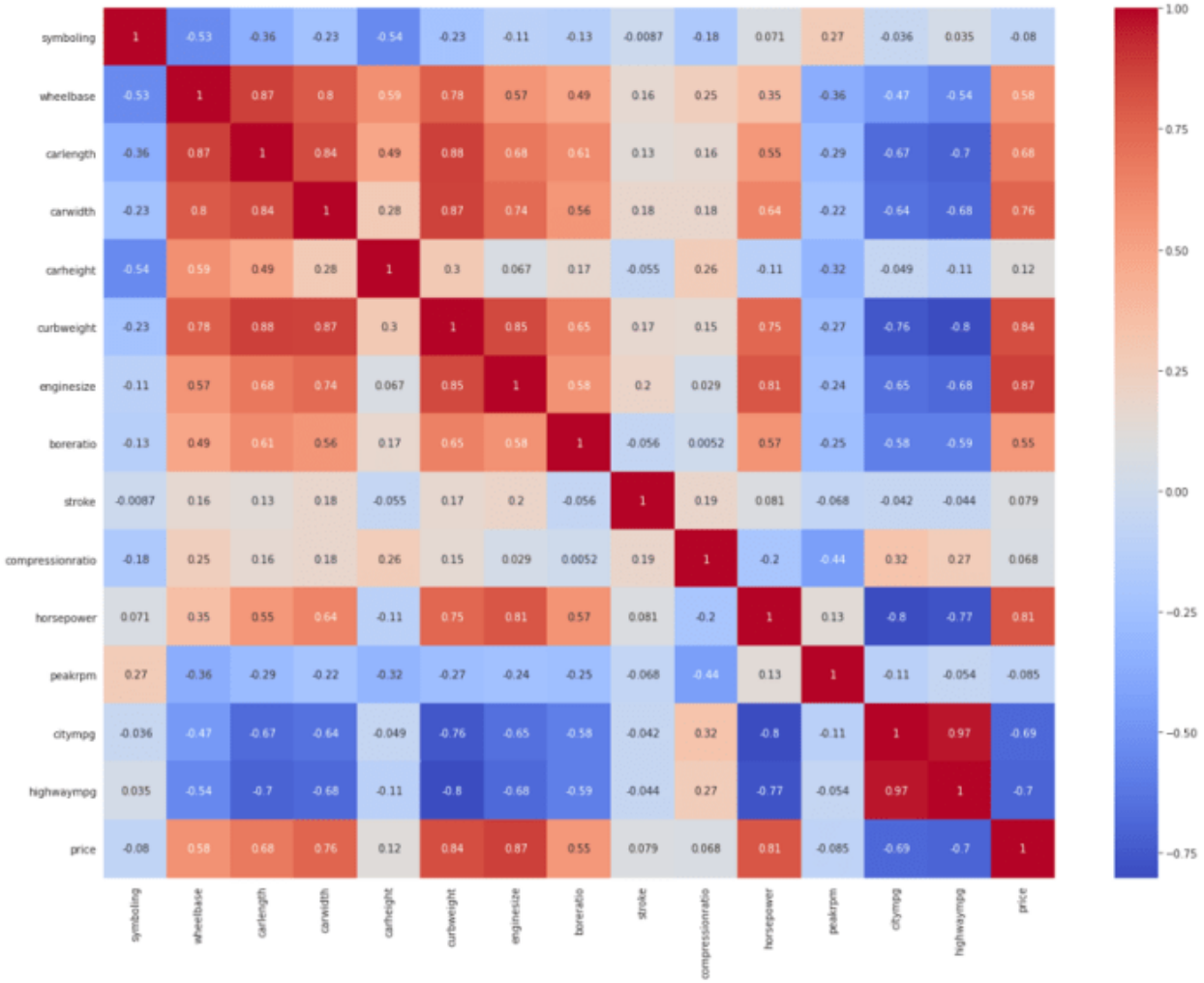
car_ID 0
symboling 0
CarName 0
fueltype 0
aspiration 0
doornumber 0
carbody 0
drivewheel 0
enginelocation 0
wheelbase 0
carlength 0
carwidth 0
carheight 0
curbweight 0
enginetype 0
cylindernumber 0
enginesize 0
fuelsystem 0
boreratio 0
stroke 0
compressionratio 0
horsepower 0
peakrpm 0

2.3 FEATURES

print(data.corr())

```
car_ID  symboling  ...  highwaympg     price
car_ID            1.000000  -0.151621  ...    0.011255 -0.109093
symboling        -0.151621   1.000000  ...    0.034606 -0.079978
wheelbase         0.129729  -0.531954  ...   -0.544082  0.577816
carlength         0.170636  -0.357612  ...   -0.704662  0.682920
carwidth          0.052387  -0.232919  ...   -0.677218  0.759325
carheight         0.255960  -0.541038  ...   -0.107358  0.119336
curbweight        0.071962  -0.227691  ...   -0.797465  0.835305
enginesize       -0.033930  -0.105790  ...   -0.677470  0.874145
boreratio         0.260064  -0.130051  ...   -0.587012  0.553173
stroke           -0.160824  -0.008735  ...   -0.043931  0.079443
compressionratio  0.150276  -0.178515  ...    0.265201  0.067984
horsepower       -0.015006   0.070873  ...   -0.770544  0.808139
peakrpm          -0.203789   0.273606  ...   -0.054275 -0.085267
citympg           0.015940  -0.035823  ...    0.971337 -0.685751
highwaympg        0.011255   0.034606  ...    1.000000 -0.697599
price            -0.109093  -0.079978  ...   -0.697599  1.000000
```

```
plt.figure(figsize=(20, 15))
correlations = data.corr()
sns.heatmap(correlations, cmap="coolwarm", annot=True)
plt.show()
```



CHAPTER 3

3. System Design:

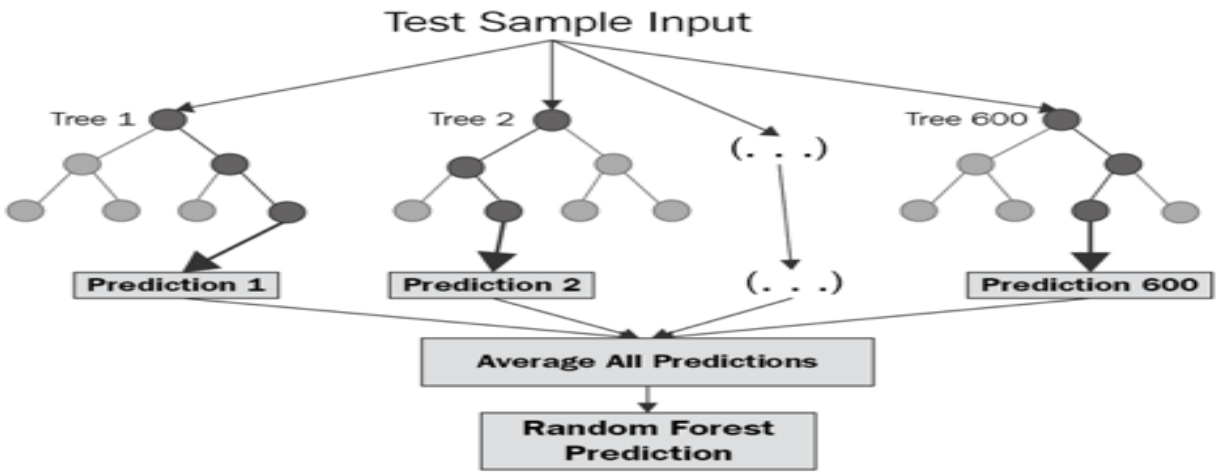
3.1 FORM DESIGN:

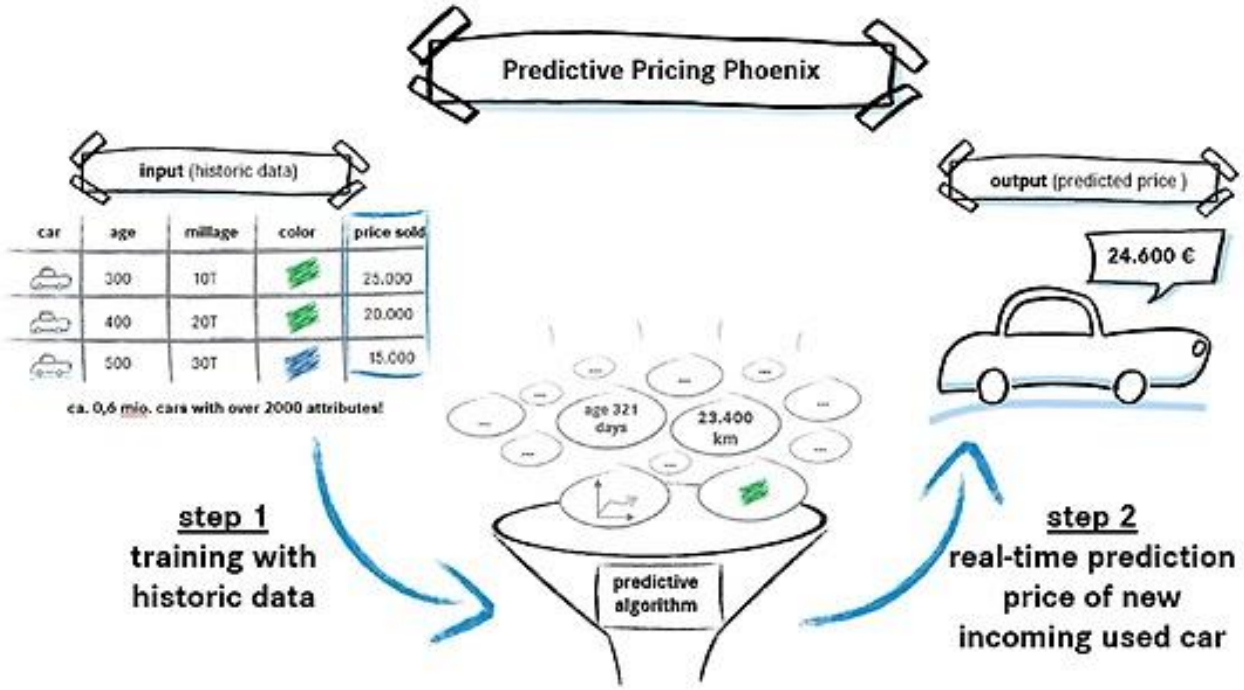
data.CarName.unique()

array(['alfa-romero giulia', 'alfa-romero stelvio',
'alfa-romero Quadrifoglio', 'audi 100 ls', 'audi 100ls',
'audi fox', 'audi 5000', 'audi 4000', 'audi 5000s (diesel)',
'bmw 320i', 'bmw x1', 'bmw x3', 'bmw z4', 'bmw x4', 'bmw x5',
'chevrolet impala', 'chevrolet monte carlo', 'chevrolet vega 2300',
'dodge rampage', 'dodge challenger se', 'dodge d200',
'dodge monaco (sw)', 'dodge colt hardtop', 'dodge colt (sw)',
'dodge coronet custom', 'dodge dart custom',
'dodge coronet custom (sw)', 'honda civic', 'honda civic cvcc',
'honda accord cvcc', 'honda accord lx', 'honda civic 1500 gl',
'honda accord', 'honda civic 1300', 'honda prelude',
'honda civic (auto)', 'isuzu MU-X', 'isuzu D-Max ',
'isuzu D-Max V-Cross', 'jaguar xj', 'jaguar xf', 'jaguar xk',
'maxda rx3', 'maxda glc deluxe', 'mazda rx2 coupe', 'mazda rx-4',
'mazda glc deluxe', 'mazda 626', 'mazda glc', 'mazda rx-7 gs',
'mazda glc 4', 'mazda glc custom l', 'mazda glc custom',
'buick electra 225 custom', 'buick century luxus (sw)',
'buick century', 'buick skyhawk', 'buick opel isuzu deluxe',
'buick skylark', 'buick century special',
'buick regal sport coupe (turbo)', 'mercury cougar',
'mitsubishi mirage', 'mitsubishi lancer', 'mitsubishi outlander',
'mitsubishi g4', 'mitsubishi mirage g4', 'mitsubishi montero',
'mitsubishi pajero', 'Nissan versa', 'nissan gt-r', 'nissan rogue',
'nissan latio', 'nissan titan', 'nissan leaf', 'nissan juke',
'nissan note', 'nissan clipper', 'nissan nv200', 'nissan dayz',
'nissan fuga', 'nissan otti', 'nissan teana', 'nissan kicks',
'peugeot 504', 'peugeot 304', 'peugeot 504 (sw)', 'peugeot 604sl',
'peugeot 505s turbo diesel', 'plymouth fury iii',
'plymouth cricket', 'plymouth satellite custom (sw)',
'plymouth fury gran sedan', 'plymouth valiant', 'plymouth duster',
'porsche macan', 'porcshce panamera', 'porsche cayenne',
'porsche boxter', 'renault 12tl', 'renault 5 gtl', 'saab 99e',
'saab 99le', 'saab 99gle', 'subaru', 'subaru dl', 'subaru brz',
'subaru baja', 'subaru r1', 'subaru r2', 'subaru trezia',

'subaru tribeca', 'toyota corona mark ii', 'toyota corona',
'toyota corolla 1200', 'toyota corona hardtop',
'toyota corolla 1600 (sw)', 'toyota carina', 'toyota mark ii',
'toyota corolla', 'toyota corolla liftback',
'toyota celica gt liftback', 'toyota corolla tercel',
'toyota corona liftback', 'toyota starlet', 'toyota tercel',
'toyota cressida', 'toyota celica gt', 'toyouta tercel',
'vokswagen rabbit', 'volkswagen 1131 deluxe sedan',
'volkswagen model 111', 'volkswagen type 3', 'volkswagen 411 (sw)',
'volkswagen super beetle', 'volkswagen dasher', 'vw dasher',
'vw rabbit', 'volkswagen rabbit', 'volkswagen rabbit custom',
'volvo 145e (sw)', 'volvo 144ea', 'volvo 244dl', 'volvo 245',
'volvo 264gl', 'volvo diesel', 'volvo 246'], dtype=object)

3.2 Input Design





Evaluating vehicles through machine learning: „Phoenix Pricing“ allows our branches to predict a used vehicle’s price precisely.

The complexity of the used car market has been growing for years and thus also the valuation of vehicles in the used car market. Besides the increasing number of model series, other factors including the selection of vehicle features and large number of suppliers make it difficult for dealers to quickly and accurately determine the value of cars. A key success criterion in the used car business is determining a "good price offer".

To support pricing experts in Mercedes-Benz business as best as possible in determining the offer price, the "Phoenix Pricing" project was developed under the management of the Controlling and IT units of Mercedes-Benz Sales Germany. With the help of machine learning, a proposed price is calculated, which delivers the most accurate prediction of the actual sales price on the basis of diverse data

This data can be obvious pricing-relevant attributes such as vehicle age and mileage, as well as special features and sales prices from previous years. The more data sources are available, the more precise the predicted price, which is continuously optimized through constant adjustment by the algorithm and integration of other data sources.

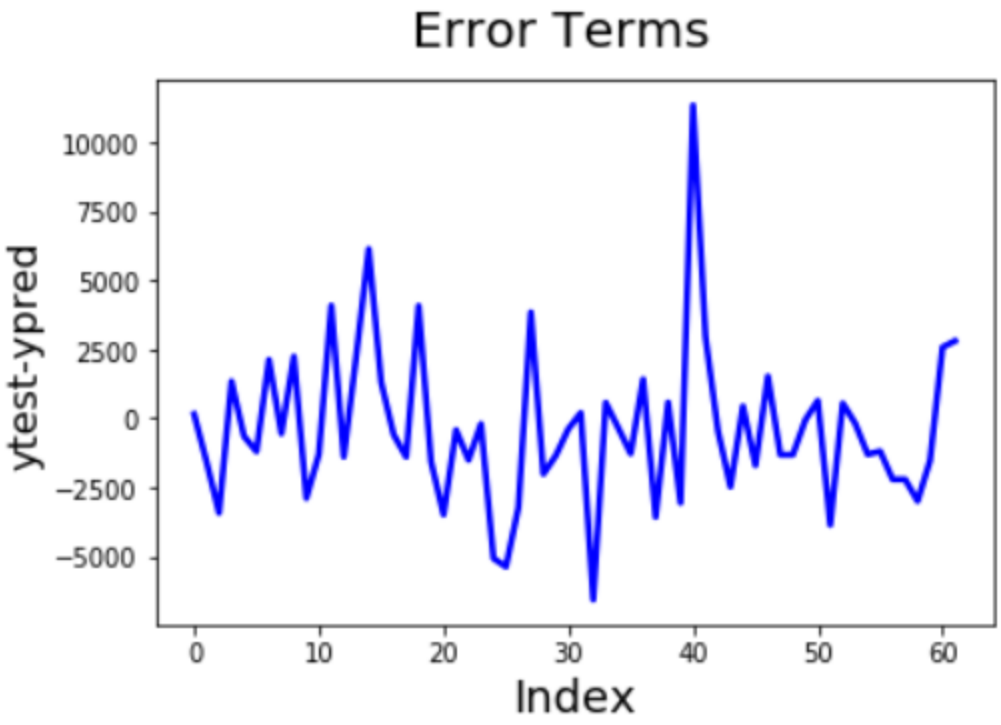
3.3 Output Design

```
y_pred=np.exp(model_rf.predict(X_test_scaled))
number_of_observations=20
x_ax = range(len(y_test[:number_of_observations]))
plt.figure(figsize=(20,10))
plt.plot(x_ax, y_test[:number_of_observations], label="True")
plt.plot(x_ax, y_pred[:number_of_observations], label="Predicted")
plt.title("Car Price - True vs Predicted data")
plt.xlabel('Observation Number')
plt.ylabel('Price')
plt.xticks(np.arange(number_of_observations))
plt.legend()
plt.grid()
plt.show()
```

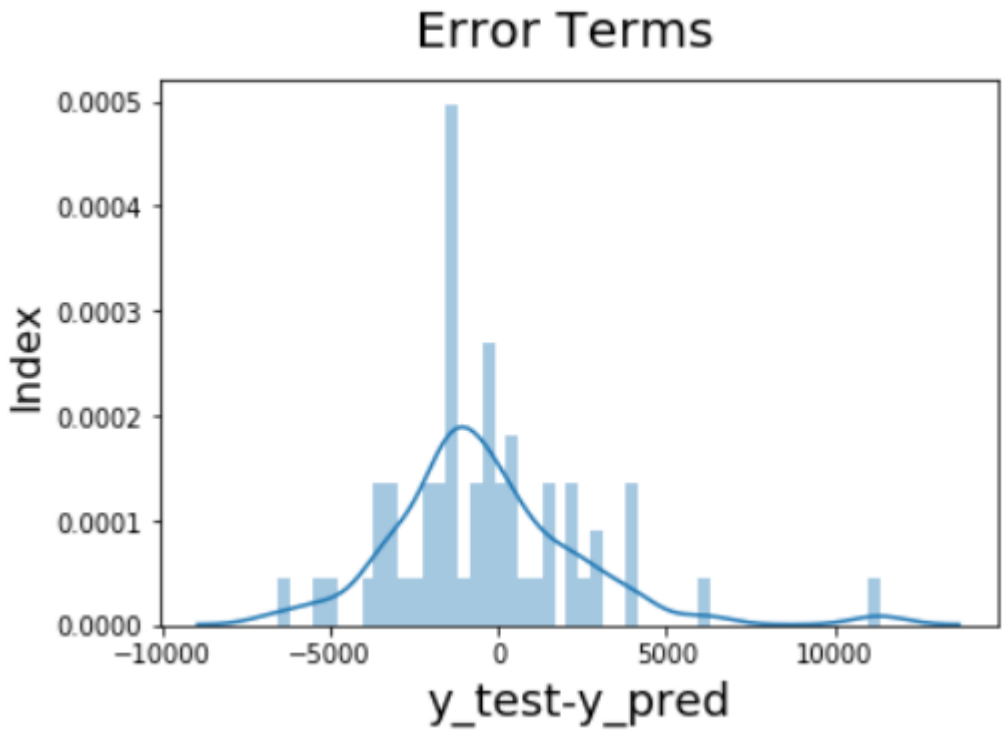
code:

```
# Loading Libraries
import pandas as pd
import numpy as np
from sklearn.preprocessing import OrdinalEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import
mean_squared_log_error,mean_squared_error,mean_absolute_error,mean_absolute
_percentage_error
import datetime
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LinearRegression
from xgboost import XGBRegressor
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns
from keras.models import Sequential
```

3.4 Database Design




```
fig = plt.figure()
sns.distplot((y_test-y_pred),bins=50)
fig.suptitle('Error Terms', fontsize=20)                # Plot heading
plt.xlabel('y_test-y_pred', fontsize=18)                # X-label
plt.ylabel('Index', fontsize=16)                       # Y-label
plt.show()
```



Now it may look like that the mean is not 0, though compared to the scale of 'Price', -380 is not such a big number (see distribution below).

Conclusion

In this article, we tried predicting the car price using the various parameters that were provided in the data about the car. We build machine learning and deep learning models to predict car prices and saw that machine learning-based models performed well at this data than deep learning-based models.

Bibliography

1.https://www.analyticsvidhya.com/blog/2021/07/car-price-prediction-machine-learning-vs-deep-learning/

2.https://group.mercedesbenz.com/careers/about-us/artificial-intelligence/for-nerds/pricing.html

3.https://www.analyticsvidhya.com/blog/2021/05/build-and-deploy-a-car-price-prediction-system/

4. https://towardsdatascience.com/predicting-car-price-using-machine-learning-8d2df3898f16

5. https://medium.com/analytics-vidhya/car-price-prediction-end-toend-machine-learning-web-application-8e9e1fcbd8b3

6. https://towardsdatascience.com/used-car-price-prediction-using-machine-learning-e3be02d977b2

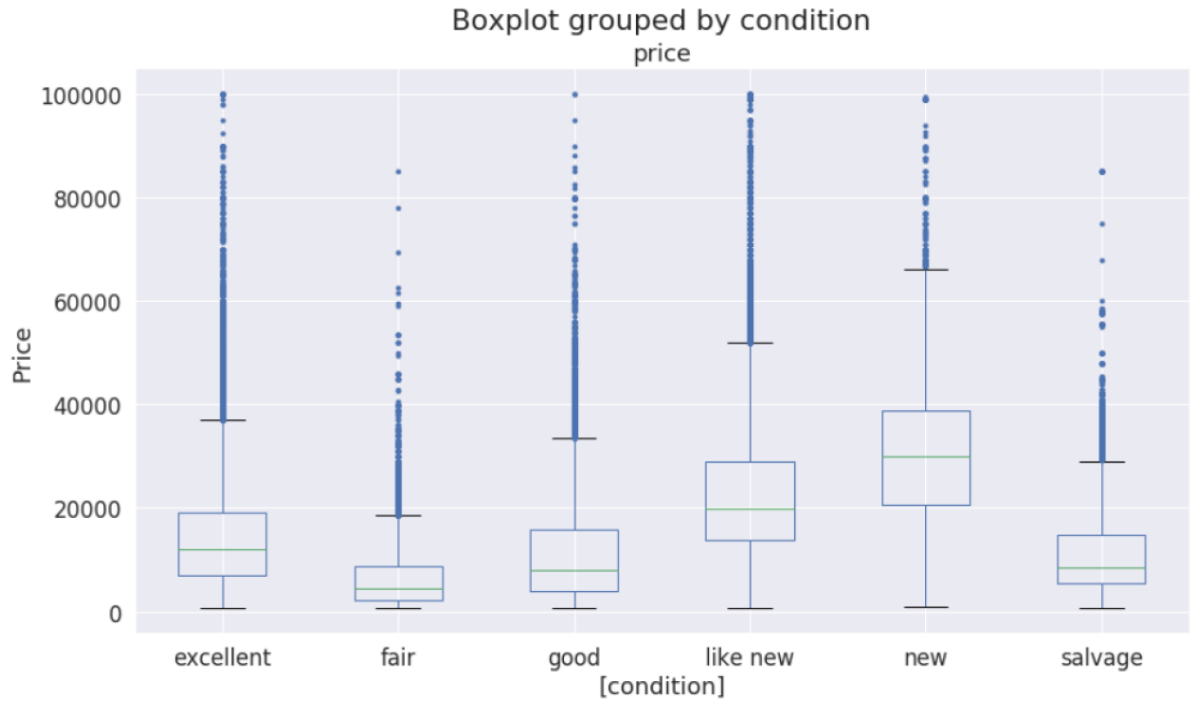
7.https://www.kaggle.com/ashydv/car-price-prediction-linear-regression

8. https://www.researchgate.net/publication/356756110_Price_Prediction_for_Pre-Owned_Cars_Using_Ensemble_Machine_Learning_Techniques

9. https://ieeexplore.ieee.org/document/9696839/

Appendices:

A.DATA FLOW DIAGRAM:



B.TABLE STRUCTURE:

#	MAKE	CYLINDER VOLUME (CC)	YEAR	MILEAGE/KM	PRICE (RS)
1	TOYOTA	1300	2007	38000	410000
2	NISSAN	1500	2007	50000	325000
3	HONDA	1500	2005	59000	385000
4	TOYOTA	1000	2007	59000	360000
5	TOYOTA	1300	1989	62665	50000
6	TOYOTA	1500	2008	67000	615000
7	TOYOTA	1500	2008	69000	575000
8	TOYOTA	1490	2006	73000	450000
9	TOYOTA	1600	2006	82000	550000
10	TOYOTA	1000	2006	85000	325000
11	TOYOTA	1500	2000	113000	325000
12	TOYOTA	1500	2000	129000	218000
13	NISSAN	1500	2001	145000	195000

C.SOURCE CODE

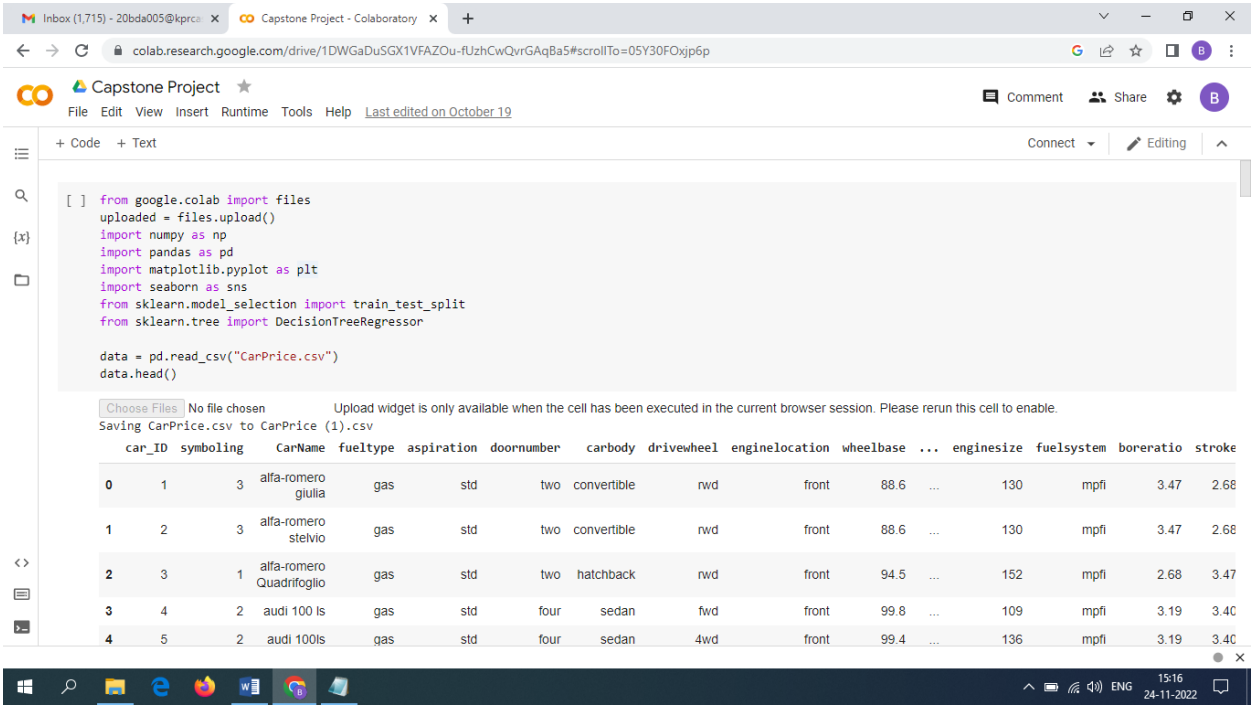
```
from google.colab import files
uploaded = files.upload()
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor

data = pd.read_csv("CarPrice.csv")
data.head()
data.isnull().sum()
data.info()
print(data.describe())
data.CarName.unique()
sns.set_style("whitegrid")
plt.figure(figsize=(15, 10))
sns.distplot(data.price)
plt.show()
print(data.corr())
plt.figure(figsize=(20, 15))
```

```
correlations = data.corr()
sns.heatmap(correlations, cmap="coolwarm", annot=True)
plt.show()
predict = "price"
data = data[["symboling", "wheelbase", "carlength",
        "carwidth", "carheight", "curbweight",
        "enginesize", "boreratio", "stroke",
        "compressionratio", "horsepower", "peakrpm",
        "citympg", "highwaympg", "price"]]
x = np.array(data.drop([predict], 1))
y = np.array(data[predict])

from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2)
from sklearn.tree import DecisionTreeRegressor
model = DecisionTreeRegressor()
model.fit(xtrain, ytrain)
predictions = model.predict(xtest)
from sklearn.metrics import mean_absolute_error
model.score(xtest, predictions)
```

D.OUTPUT



Capstone Project

File Edit View Insert Runtime Tools Help Last edited on October 19

Comment Share

+ Code + Text

Connect Editing

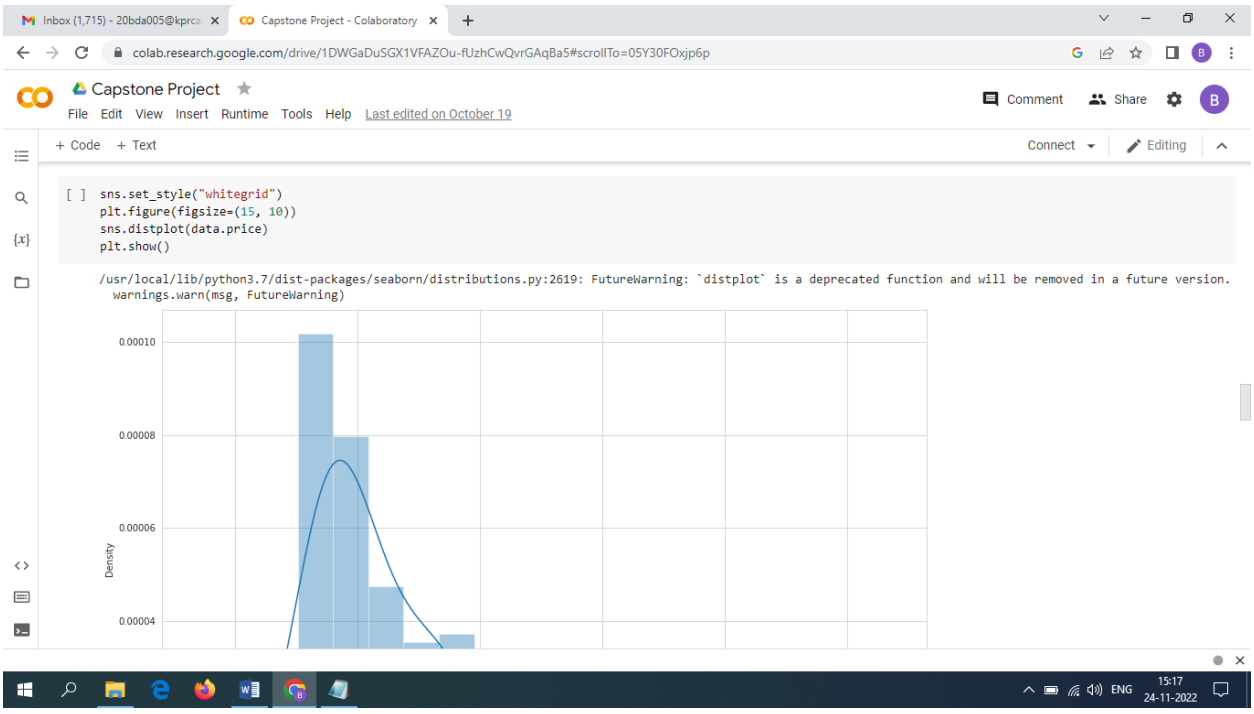
```
[ ] from google.colab import files
    uploaded = files.upload()
    import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
    from sklearn.model_selection import train_test_split
    from sklearn.tree import DecisionTreeRegressor

    data = pd.read_csv("CarPrice.csv")
    data.head()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving CarPrice.csv to CarPrice (1).csv

	car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	...	enginesize	fuelsystem	boreratio	stroke
0	1	3	alfa-romero giulia	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68
1	2	3	alfa-romero stelvio	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68
2	3	1	alfa-romero Quadrifoglio	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47
3	4	2	audi 100 ls	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.40
4	5	2	audi 100ls	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.40



Capstone Project

File Edit View Insert Runtime Tools Help Last edited on October 19

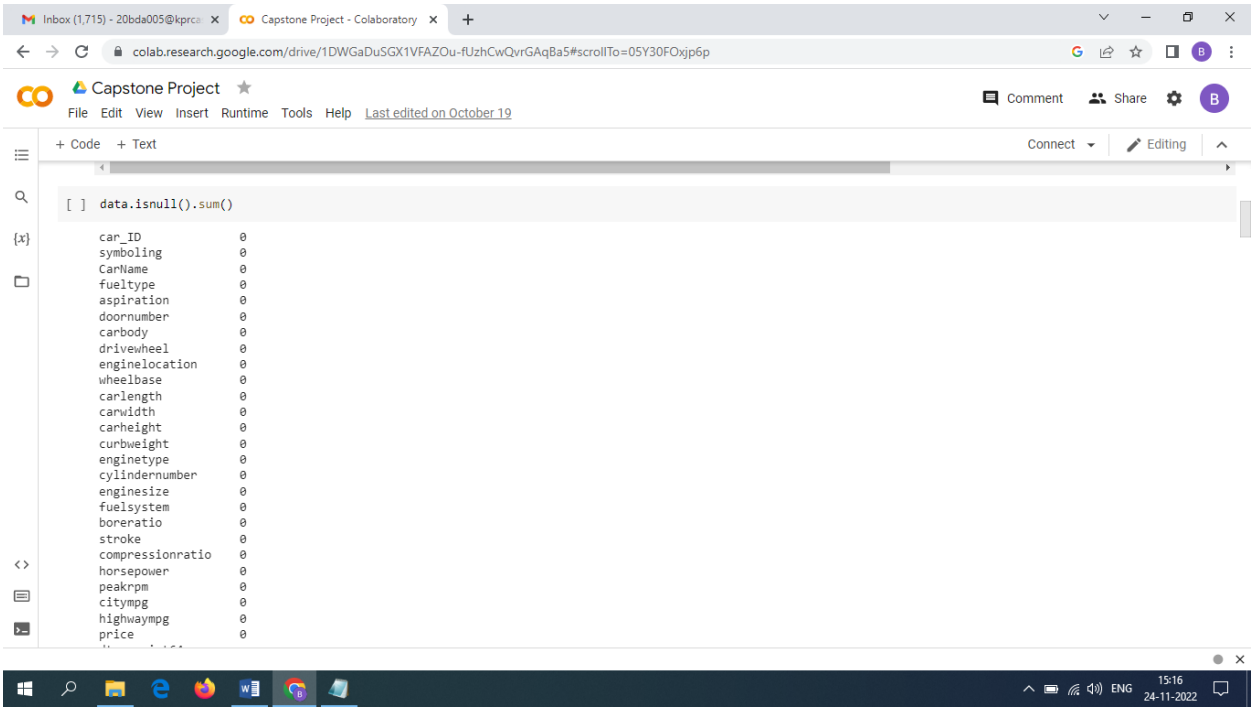
Comment Share

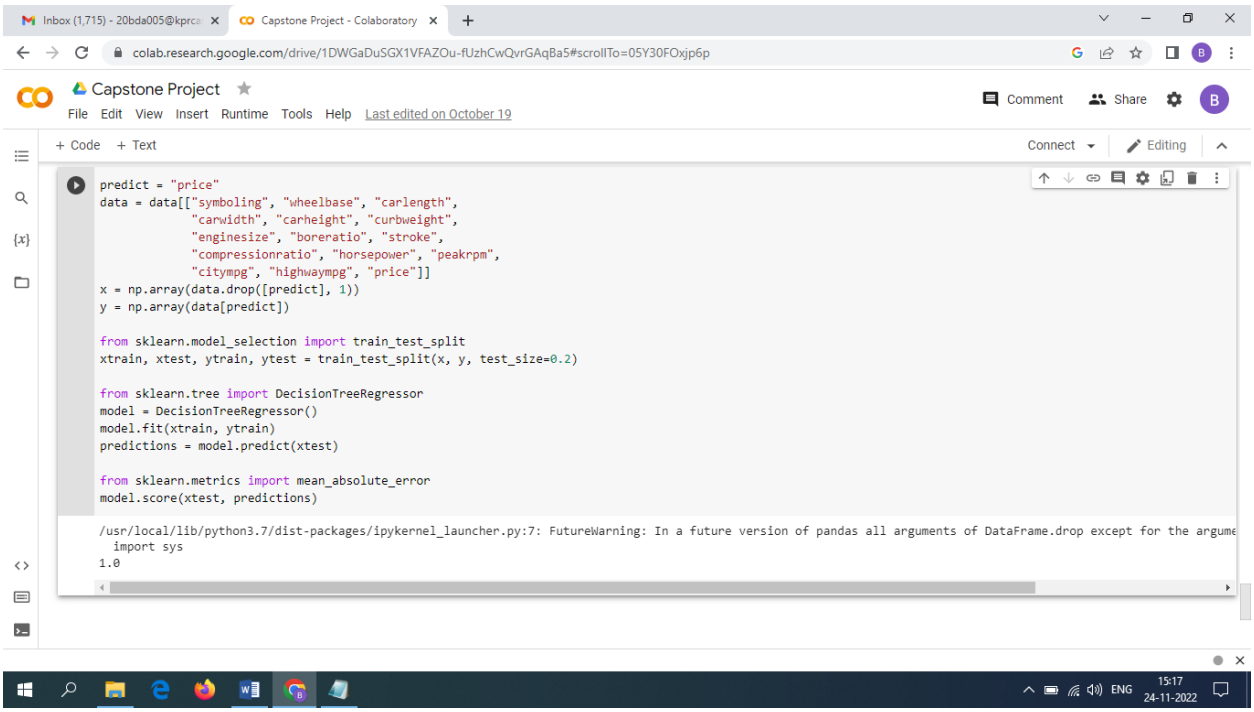
+ Code + Text

Connect Editing

```
[ ] sns.set_style("whitegrid")
    plt.figure(figsize=(15, 10))
    sns.distplot(data.price)
    plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version.
  warnings.warn(msg, FutureWarning)
```


Capstone Project - Colaboratory
colab.research.google.com/drive/1DWGaDuSGX1VFAZOu-fUzhCwQvrGAqBa5#scrollTo=05Y30FOxjp6p
Capstone Project
File Edit View Insert Runtime Tools Help Last edited on October 19
Comment
Share
+ Code + Text
Connect
Editing
[ ] data.isnull().sum()
car_ID 0
symboling 0
CarName 0
fueltype 0
aspiration 0
doornumber 0
carbody 0
drivewheel 0
enginelocation 0
wheelbase 0
carlength 0
carwidth 0
carheight 0
curbweight 0
enginetype 0
cylindernumber 0
enginesize 0
fuelsystem 0
boreratio 0
stroke 0
compressionratio 0
horsepower 0
peakrpm 0
citympg 0
highwaympg 0
price 0
ENG
15:16
24-11-2022


Capstone Project - Colaboratory
colab.research.google.com/drive/1DWGaDuSGX1VFAZOu-fUzhCwQvrGAqBa5#scrollTo=05Y30FOxjp6p
Capstone Project
File Edit View Insert Runtime Tools Help Last edited on October 19
Comment
Share
+ Code + Text
Connect
Editing
predict = "price"
data = data[["symboling", "wheelbase", "carlength",
"carwidth", "carheight", "curbweight",
"enginesize", "boreratio", "stroke",
"compressionratio", "horsepower", "peakrpm",
"citympg", "highwaympg", "price"]]
x = np.array(data.drop([predict], 1))
y = np.array(data[predict])
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2)
from sklearn.tree import DecisionTreeRegressor
model = DecisionTreeRegressor()
model.fit(xtrain, ytrain)
predictions = model.predict(xtest)
from sklearn.metrics import mean_absolute_error
model.score(xtest, predictions)
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:7: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argume
import sys
1.0
ENG
15:17
24-11-2022