AUTOMOBILE PRICE PREDICTOR

IC SOLUTIONS

INSTRUCTOR- Abhishek . C

student name- Prajodh Pragath

USN Number-1BY19CS104

Email id- prajodhpragaths@gmail.com

Acknowledgement

Would like to express my special thanks of gratitude to Abhishek sir our course instructor for his guidance and support in completeing this project .I would also like to extend my gratitude to IC Solutions and takeiteasy engineers for making this course possible and giving me the opportunity to do this project.

Secondly I would also like to thank my parents and friends who helped me a lot in finalizing this project within the given time frame

ABSTRACT

A car prediction project can be of great help to companies starting out in a different region to get an idea about the public demand. Automobile price predictor is an project using available ML algorithms to predict the price of the cars based on its attributes. The type of machine learning models which can be used are regression models as we are predicting a value. In this project the ml algorithms used were linear regression, random forests and bagging regressor. The data used was provided by our course instructor which is data obtained from the region of USA.

The three algorithms were tested and the one with the best r2 score was chosen which was the bagging regressor. Which is a ensemble meta-estimators that fits base regressors each on random subsets of the original dataset and accumulates all the predictions to form an final value. It had an r2 score of 0.90351429 which is 90.35% accurate

ABOUT THE COMPANIES

ICS is a digital service provider that aims to provide software, designing and marketing solutions to individuals and businesses. At ICS, we believe that service and quality is the key to success.

We provide all kinds of technological and designing solutions from Billing Software to Web Designs or any custom demand that you may have. Experience the service like none other!

Some of our services include:

Development - We develop responsive, functional and super fast websites. We keep User Experience in mind while creating websites. A website should load quickly and should be accessible even on a small view-port and slow internet connection.

Mobile Application - We offer a wide range of professional Android, iOS & Hybrid app development services for our global clients, from a startup to a large enterprise.

Design - We offer professional Graphic design, Brochure design & Logo design. We are experts in crafting visual content to convey the right message to the customers.

Consultancy - We are here to provide you with expert advice on your design and development requirement.

Videos - We create a polished professional video that impresses your audience.

INDEX

INTRODUCTION

This internship was titled Machine learning in python. It was an ten day internship for engineering students to get started with machine learning and AI. The scope of the internship and things which were covered during the course were Introduction to AI, Python Essentials, Data Analysis Libraries like numpy and pandas, Data Visualization Libraries like matplotlib and seaborn. Then we went through ML models like Linear Regression, Logistic Regression, Support Vector Machines, Decision Trees, Random Forest and K Nearest Neighbours. We were also given a brief introduction to deep learning and the course ended with Model Deployment.

PROBLEM STATEMENT AND OBJECTIVE

Problem statement-Every project starts with a problem even mine did. Cars are coming out with more and more features day by day. Some of these features cost more and some cost less. A list of these attributes of a car can make it possible to predict its price but doing this manually isn't feasible and might consume a lot of man power. Hence my project with the help of Machine learning makes it easy and possible to predict the prices of cars based on their features and attributes by providing it with the data set of a group of cars with their prices and features

Objective- Our goal is to create a fully functioning and effective program with an ML model which will help us with prediction of the prices of the car. First the dataset set should be loaded and analysed . then we have to do the data cleaning process to make sure the data set fits the model then we check the accuracy of our model to check if it is effective in predicting the price

REQUIREMENT SPECIFICATION

Hardware requirements-

The system should full fill these following hardware requirements

- processor with a clock speed of atleast 2 GHz eg-i3 intel

- 4 GB ram

- hard drive storage of minimum 64 GB

-wifi or ethernet connection for internet

Software requirements-

The system should full fill these following software requirements

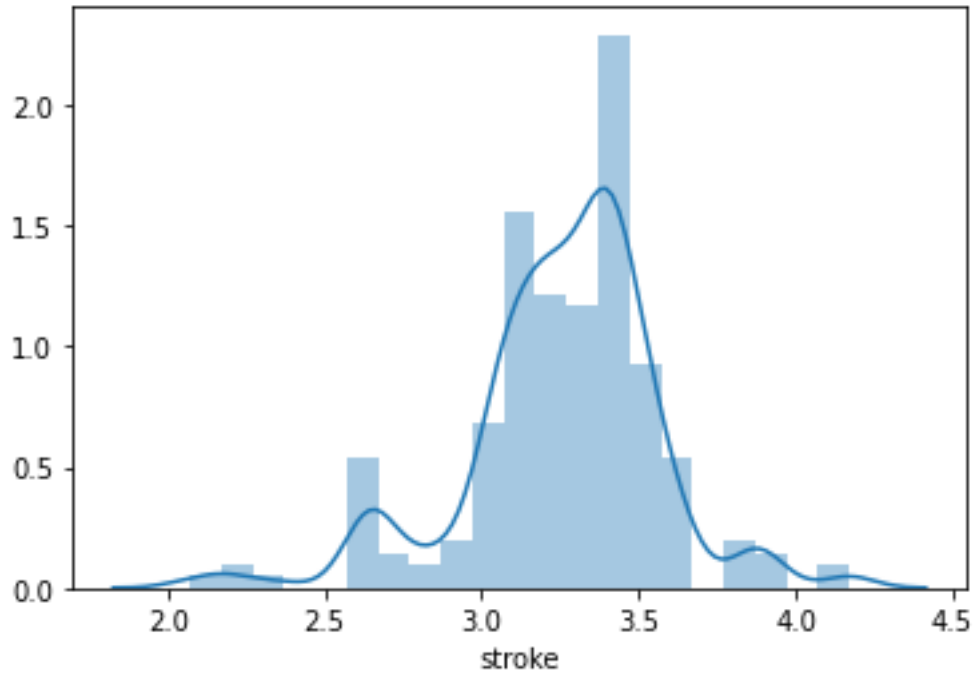
- OS can be windows xp or later
- Google chrome
- Python 3
- Jupyter notebook
- Anaconda prompt

EXPLORATORY DATA ANALYSIS

Code-

```
sns.distplot(newauto['stroke'])
```

Graph



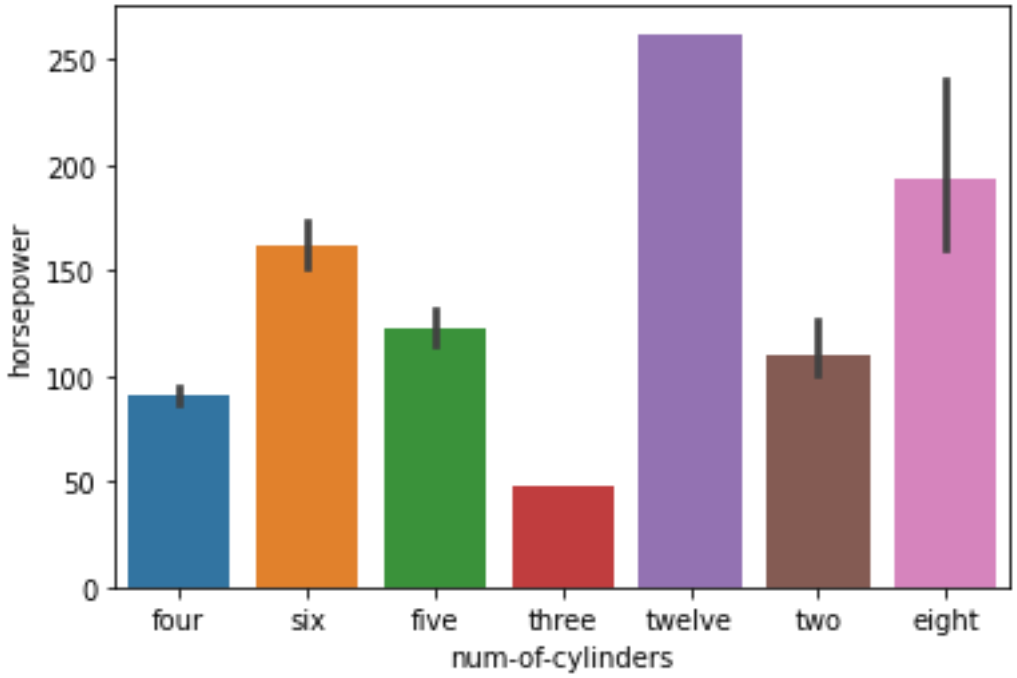
Insights-

most of the cars have 3 to 3.5 stroke

Code –

```
sns.barplot(x='num-of-cylinders',y='horsepower',data=newauto)
```

Graph-



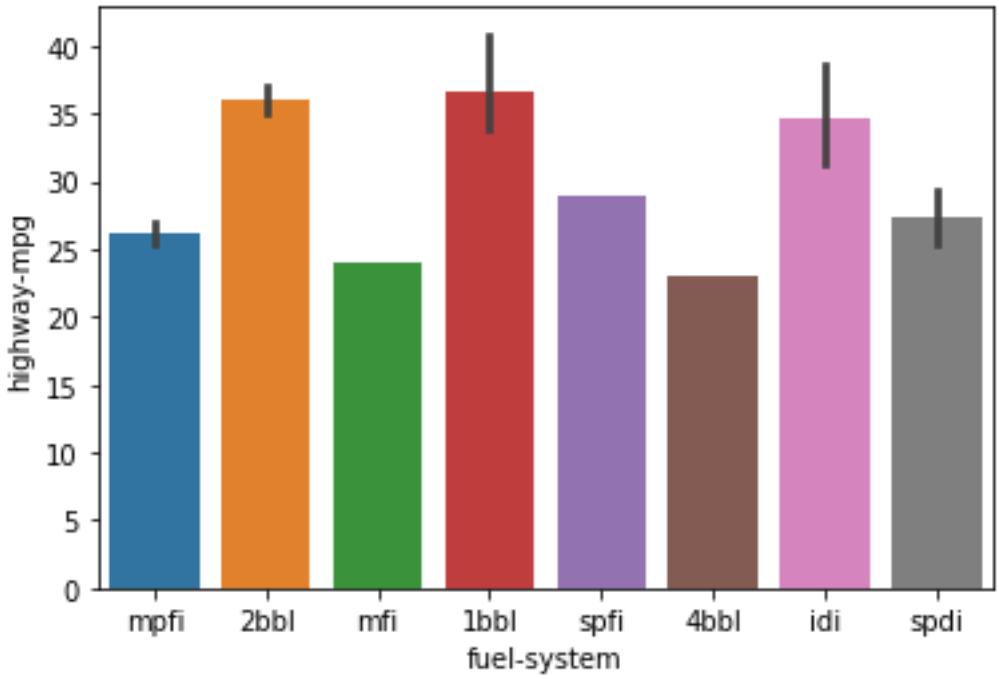
Insight-

higher the no of cylinders higher is the horsepower of the car

Code-

```
sns.barplot(x='fuel-system',y='highway-mpg',data=newauto)
```

Graph-



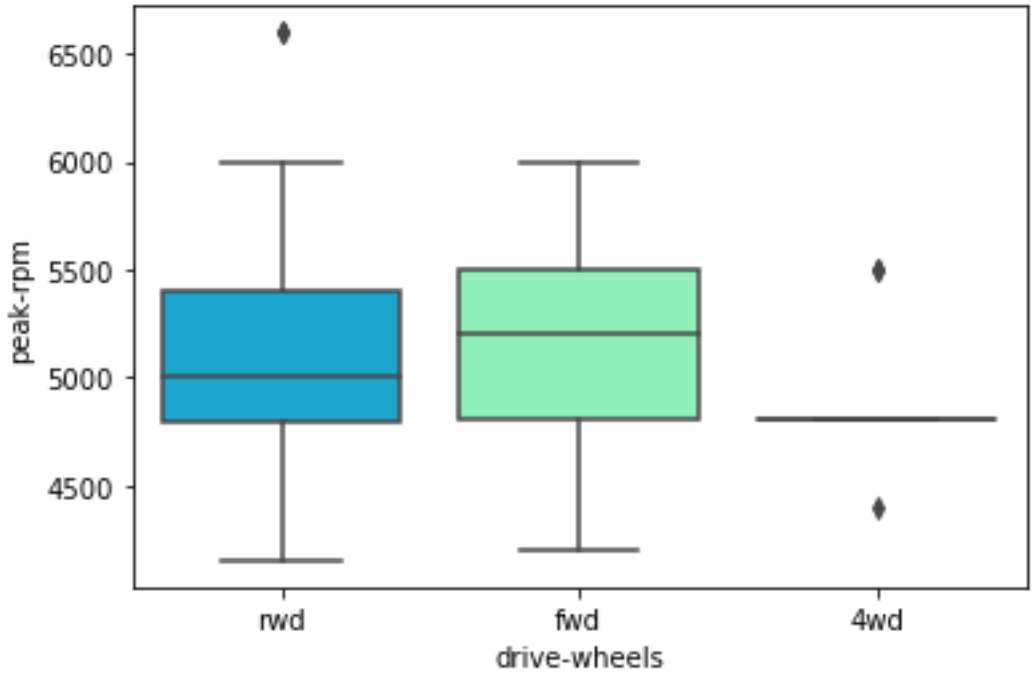
Insights

the 2bbl,1bbl and idi fuel systems are best suited for higher milages performances

Code-

```
sns.boxplot(x="drive-wheels", y="peak-rpm", data=newauto,palette='rainbow')
```

Graph-



Insights

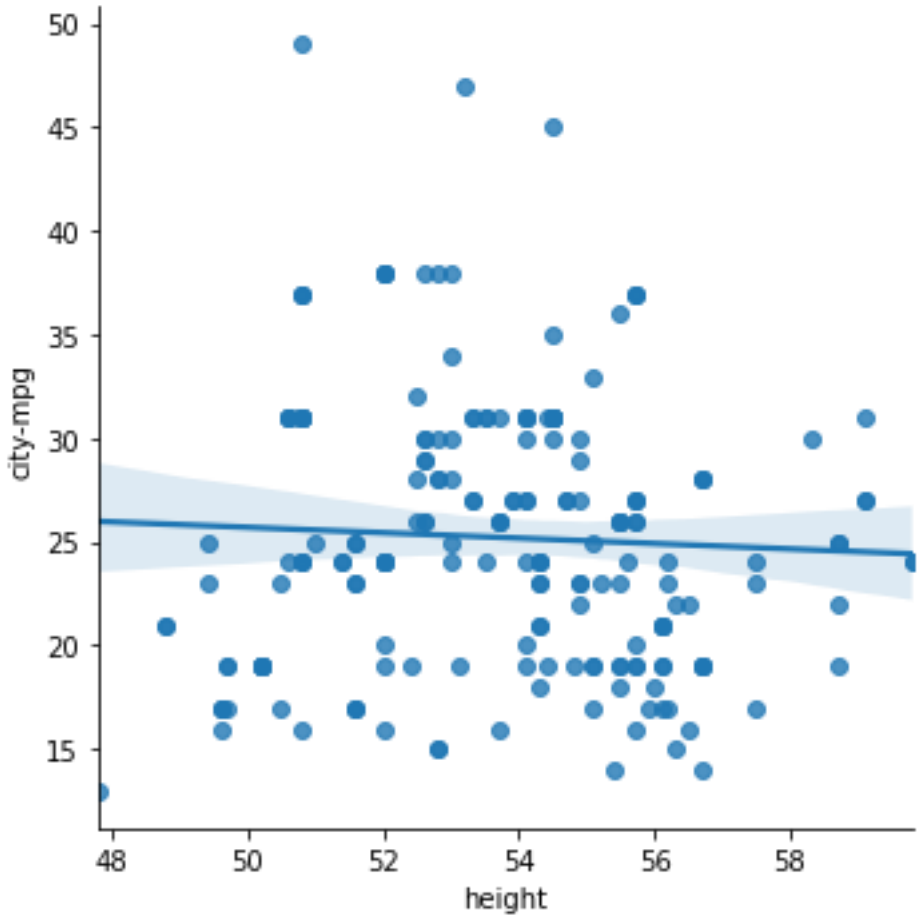
-The fwd has a larger peakrpm and peakrpm range

-fwd has the lowest peakrpm and peakrpmrange

Code

```
sns.lmplot(x="height", y="city-mpg", data=newauto)
```

Graph



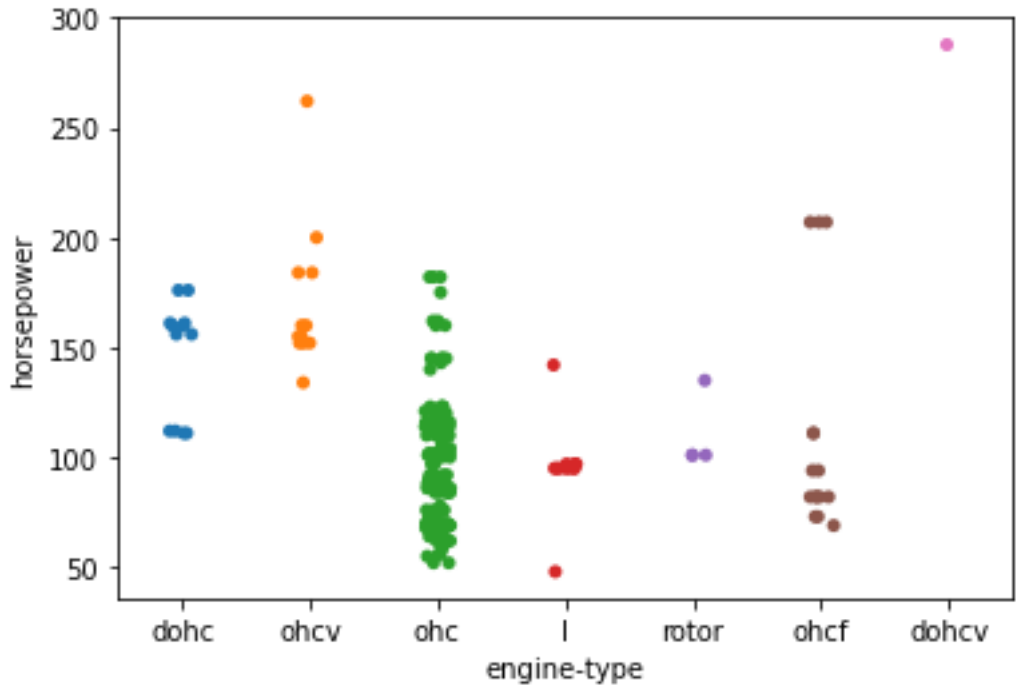
Insights

cars with height ranging from 52 to54 have an average city mpg of 20 to 25

Code

```
sns.stripplot(x="engine-type", y="horsepower", data=newauto,jitter=True)
```

Graph



Insights

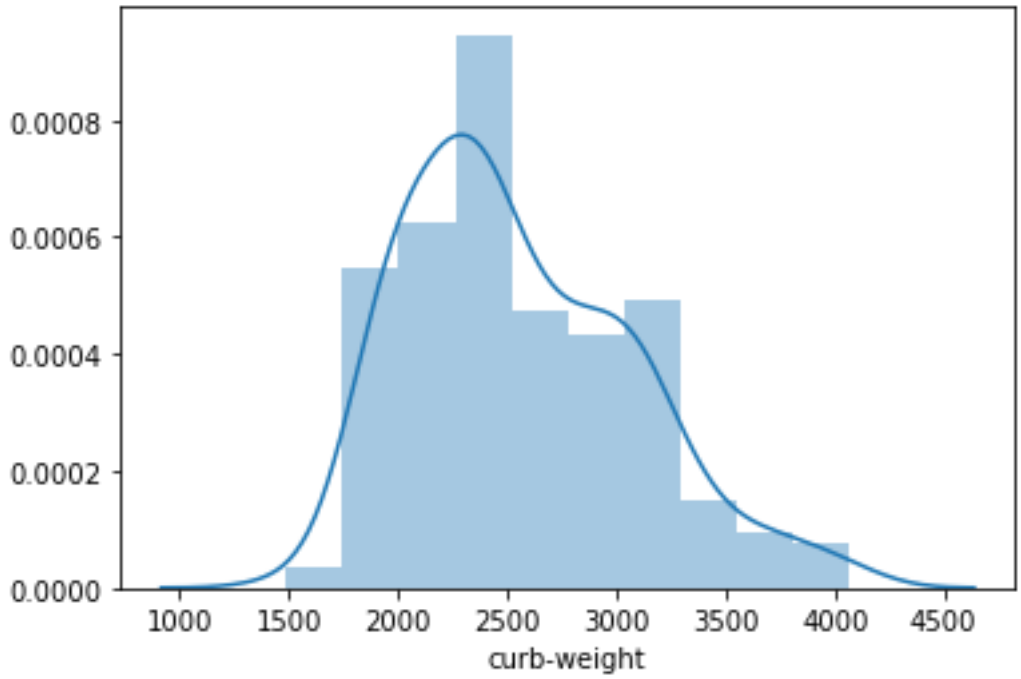
-most of the cars belong to ohc category of engine type

-ohvc engine type usually has a higher horsepower range

Code

```
sns.distplot(newauto['curb-weight'])
```

Graph



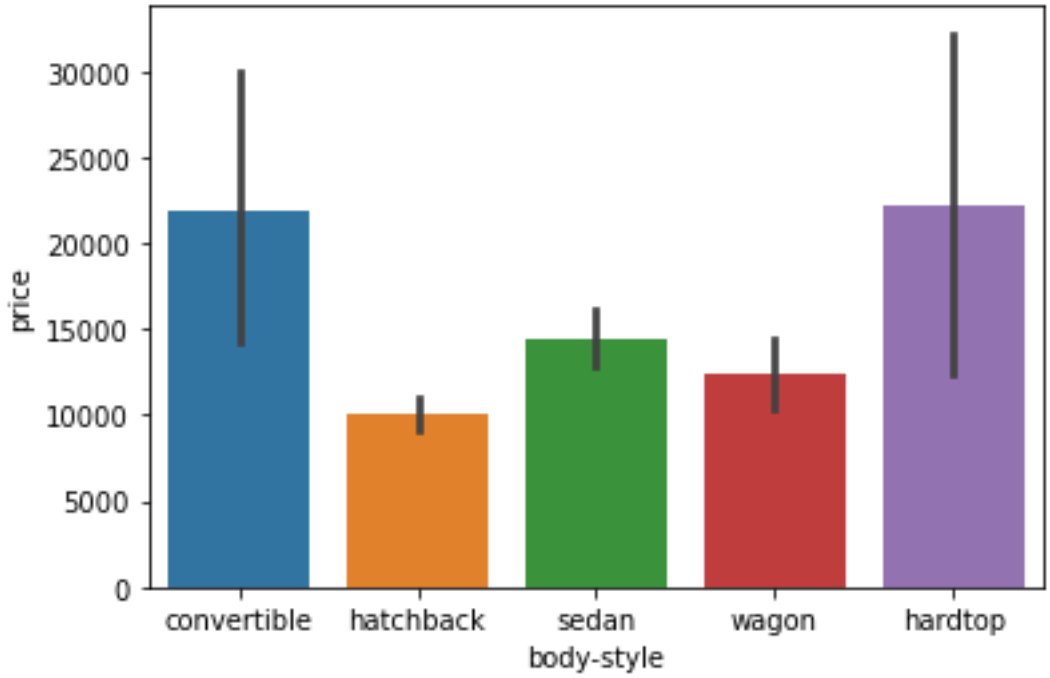
Insights

most of the the cars have an curb weight ranging from 1750 to 3250

Code

```
sns.barplot(x='body-style',y='price',data=newauto)
```

Graphs



Insights

-hardtop and convertibles are more expensive that the other bodystyles

-hatchback and wagon are the cheaper bodystyles

PREPARING MACHINE LEARNING MODEL

This project requires the regression model to be used as we have to predict a numerical value the price of the car. Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent (target) and independent variables.

PREPARING THE DATA FOR TRAINING

Code-

```
y=newauto['price']

X=newauto.drop(['price'],axis=1)

from sklearn.model_selection import train_test_split

X_train,X_test,y_train,y_test=train_test_split(X,y,random_state=101,test_size=0.3)
```

Linear regression

Code

```
from sklearn.linear_model import LinearRegression

from sklearn.metrics import r2_score

Lr=LinearRegression()

Lr.fit(X_train,y_train)

pred1=lr.predict(X_test)

score_lr=r2_score(y_test,pred_1)
```

Output

R2_score

```
0.8021470958755839
```

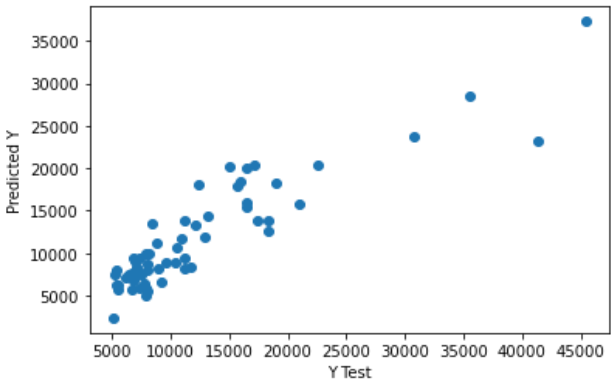
Code

```
plt.scatter(y_test,pred)

plt.xlabel('Y Test')

plt.ylabel('Predicted Y')
```

Output



Random forest trees

Code

```
from sklearn.ensemble import RandomForestRegressor

rfg=RandomForestRegressor()

rfg.fit(X_train,y_train)

pred2=rfg.predict(X_test)

score_rfr=r2_score(y_test,pred2)
```

output

R2_score

```
0.9028826563791924
```

Bagging regressor

Code

```
from sklearn.ensemble import BaggingRegressor

br=BaggingRegressor()

br.fit(X_train,y_train)

pred3=br.predict(X_test)

score_br=r2_score(y_test,pred3)
```

output

R2_score

```
0.9035142934949485
```

ML model chart

Serial number	Algorithm name	R2_score
1	Linear regression	0.8021470958755839
2	Random forest	0.9028826563791924
3	Bagging regressor	0.9035142934949485

HURDLES

Hurdles is a word associated with expressing the difficulties faced which doing a project or any work. This project to had a obstacles that I had to face. The first problem I faced was while I tried to go ahead with exploratory data analysis without checking my data set which had to first undergo data cleaning. the course instructor helped me with this problem as I mailed him about this problem.

The other obstacles were majorly programming errors and coding error which mostly I was able to figure out else reached out to my friends for help.

CONCLUSION

This project on automobile price predictor is about predicting prices of cars when provided with the features of the car. This was achieved by first obtaining a data set of cars of various features and prices. Then analysing and cleaning the data was done so that we can use an suitable ML models to do the price predicting. Three ML models were used and the r2 score were compared to see how accurate . the most accurate model in my project was the bagging regressor which had a r2 score of 0.90351.

This project majorly helped to boost my confidence on the basis of project development. It helped me to apply the concepts we learned during the course which makes it easier to remember in the long run. It provoked me to think more analytically.

BIBLOGRAPHY

www.Researchgate.net

www.wikepedia.org

www.scikit-learn.org

software engineering by Ian Sommerville

python data visualization cook book by Igor Milovanović