WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Fuel Efficiency Prediction

Abhay Pratap, Abhishek Kumar, Ahamad Ali, Anup Kumar

Department of Computer Science Engineering, RKGIT

ABSTRACT:

Car manufacturers are always trying to improve fuel efficiency because the car industry has been growing for over 200 years, and fuel prices keep rising. Customers are becoming more picky about features, so car makers are constantly updating their processes to improve fuel efficiency. What if there was an accurate way to predict a car's MPG (Miles per Gallon) or fuel consumption based on certain known details? By creating a more efficient car, manufacturers could outperform competitors, increase demand, and boost production.

Using Machine Learning, we are developing prediction models to reduce error rates for cars made in recent years. We will use available datasets to create a model that predicts the fuel efficiency of various vehicles from different times. These datasets will include details like the number of cylinders, engine displacement, horsepower, and weight. Machine Learning is suitable for this analysis because it can identify patterns in the data and build models from them. Additionally, we will use deep learning techniques to create other models. The analysis will show which model has the least error and the best efficiency.

Keywords: Machine Learning, Fuel Efficiency Prediction, Random Forest, Decision tree, KNN, Linear Regression.

1. Introduction

Understanding what influences fuel consumption and being able to predict it is crucial for optimizing fuel efficiency. In the transportation industry, Miles per Gallon (MPG) is used to calculate a vehicle's efficiency based on the energy it uses. MPG varies depending on the origin of the vehicle. To check the MPG of vehicles, we have created chart models. These chart models show the MPG of vehicles based on the number of cylinders, engine displacement, horsepower, and weight. Engine size is measured by displacement, which is usually expressed in liters or cubic centimeters. The origin is a discrete number ranging from 1 to 3. Based on this dataset, we assume that 1 represents a vehicle from America, 2 represents a vehicle from Europe, and 3 represents a vehicle from Asia or other places. Some of the values in this dataset might be incorrect, so we will correct those values during data preprocessing. Modeling fuel consumption on highways is easier because external factors like traffic and road conditions do not significantly affect fuel consumption. Additionally, being able to predict fuel consumption can help owners detect potential fuel fraud if any occurs.

Manufacturers, regulators, and customers are all interested in vehicle fuel consumption models. These models are needed during all stages of a vehicle's life cycle. The goal of this work is to model the average fuel consumption for heavy vehicles throughout their operation and maintenance. Generally, there are three types of methods for creating fuel consumption models:

Empirical Models: These models are based on real-world data and observations. They use historical data to identify patterns and relationships between different factors and fuel consumption.

Material science-based models, these are the models that are framed from an exhaustive handle of the actual framework. These models utilize exhaustive numerical conditions to depict the elements of the vehicle's parts at each time step.

- Factual models, which are additionally information driven and lay out a planning between the likelihood circulation of a chose set of indicators and the objective result.
- AI models, which are information driven and address a theoretical planning from an info space comprising of a chose set of indicators to a result space that addresses the objective result, for this situation normal fuel utilization.

The choice between these techniques is determined by cost and accuracy, depending on the needs of the intended application.

Without precise information on the vehicle's actual properties and measurements, the method should be able to adapt to a wide range of vehicle technologies (including future ones) and configurations for each vehicle. While gauging the required exactness versus the expense of creating and adjusting an individualized model for every vehicle, AI arises as the strategy for decision. There have been a few past models created for both immediate and normal fuel use Since they can represent the system's movement at different time steps, physics-based models are the most suitable for evaluating short-term fuel use. Since identifying patterns in real-time data is complex, other methods may be less effective for this purpose.

troublesome, AI models can't foresee prompt fuel utilization with an elevated degree of exactness. These calculations, then again, are prepared to do precisely recognizing and learning patterns in normal fuel use Recently proposed AI methods for average fuel consumption use a set of indicators collected over time to estimate fuel use in gallons per mile or liters per kilometer. While our method still focuses on average fuel consumption, it differs from earlier models by quantizing the indicators' input based on a fixed distance instead of a fixed time period. In the proposed models, all indicators are collected within a fixed window that represents the distance traveled by the vehicle, resulting in a better mapping from the input data to the model's output. Previous AI models, on the other hand, not only had to learn the patterns in the input data but also had to convert from a time-based scale to a distance-based scale (i.e., average fuel consumption). Using the same scale for both the model's input and output areas has several advantages.

2. Related Work

In [1], the authors evaluated the predictive ability of three AI models to forecast the fuel consumption of a long-distance public bus. Some important factors, such as load, engine RPM, and traffic, were not included in the selected dataset, even though they directly affect fuel consumption. Despite missing these key factors, they showed that the RF (Random Forest) model could more accurately predict fuel consumption by identifying data patterns. An example of using such a model is detecting fuel fraud by comparing the actual fuel use of the vehicle to the predicted value based on various parameters like distance, location, elevation, speed, and day of the week. They plan to include more factors that influence fuel use, such as traffic, weather, and bus load, in future work to improve prediction accuracy further. They are also working on a module that will guide you through reengineering methods to reduce fuel use through better fleet management and driving habits.

The dataset used to build their model includes information about a specific long-distance public bus in Sri Lanka. The bus leaves from the Depot around 4:00 p.m. and heads to Colombo (the business capital). Then, it departs from Colombo at 7:00 p.m., travels along the A2, A4, and AB10 highways, and arrives at the destination around 7:00 a.m. the next day.

Altogether, the transport covers 365 kilometers in a single heading. The return trip follows similar way and happens between 4:00 p.m. what's more, 7:00 a.m. the following morning.

A rugged locale makes up about 33% of the course. The bus is equipped with a GPS-based tracking system and a high-accuracy capacitive fuel sensor. Data collected by these devices is transmitted almost in real-time to a cloud server via a 3G connection. The dataset includes both outbound and inbound trips between May 13 and August 31, 2015. Based on the characteristics and information in the dataset, the authors implemented various AI algorithms and techniques to predict the fuel consumption of the fleet vehicles on that specific route, taking into account different latitude and longitude factors. They used Random Forest, Gradient Boosting, and Artificial Neural Network models. After testing, they evaluated the performance using different error metrics, including Bias, MAE (Mean Absolute Error), and RMSE (Root Mean Squared Error).

In [2], the authors developed a model that accurately estimates a vehicle's MPG using various vehicle data. They improved the model, reducing the RMSE score from an initial 3.26 to an adjusted 1.97. This model can be updated with newer vehicle data and used to predict the R2 score. They found a steady increase in the R2 value from 0.82 to 0.91, indicating that the model is more reliable and useful for future MPG estimates for new vehicles.

Companies can focus resources on creating more efficient and popular vehicles that outperform competitors. Although their model might be inaccurate at times, they acknowledged that the dataset might have some incorrect MPG values. However, overall, the predictions are more accurate than the values in the dataset. Data collected from newer vehicles is much more reliable, so their model will work more efficiently with different, more accurate datasets.

They included data from previous years' vehicles in their database to predict fuel efficiency or miles per gallon using the best AI model they found, which was Linear Regression, after testing various algorithms. The calculations were done and displayed in a way that aimed for the RMSE error to be as efficient as possible, ideally between 1 and 2, to improve the accuracy of the predictions based on the vehicle's data.

3. METHODOLOGIES

In this project, we'll use different algorithms like Linear Regression, Random Forest, Decision Tree, and Neural Network. First, we'll clean and preprocess our data, which involves fixing any errors and filling in missing values. We'll also create visualizations, like graphs and plots, to explore the data. Cleaning the data means making sure it's accurate and complete before using it for analysis.

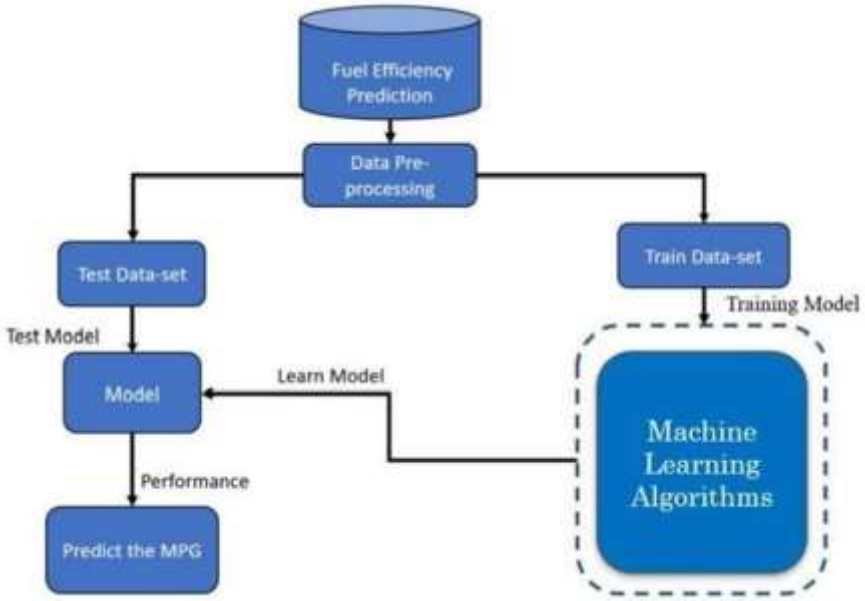
Then, we'll build different models using various algorithms and begin training them with the data. Each model will be constructed using a different algorithm. The predictions made by these models will rely on the proper training of the data using information from the dataset. Finally, we'll test the models on the datasets to evaluate their performance.

The RMSE values of all the models will be compared in order to find the better suited and efficient model. Once the models are finalized, the deployment of it can be done.

Deployment process includes the creation of a web page with the suitable models and deploys them in that web page. The user either from the automobile industry side or the customer will visit the page or give the new inputs as per their models to be checked.

The given input will be fed to the model and predictions of the fuel efficiency will be calculated accordingly. Finally, the predicted output which will be the MPG or fuel efficiency will be displayed on to the screen.

FLOWCHART



4. DATA PREPROCESSING

As a data set (also known as a dataset) is a collection of information. A data set correlates to one or more database tables in the case of tabular data, where each column of a table represents a specific variable and each row corresponds to a specific record of the data set in the requirements. The dataset that we have used in our project consists of 3032 rows and 8 columns. This dataset is taken from Kaggle and combined it with 100,000 used cars dataset. We have chosen our dataset with respect to the requirement of the project. The following consists of sample data from our dataset.

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	car name
0	18.0	8	307.0	130	3504	12.0	1970	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	1970	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	1970	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	1970	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	1970	1	ford torino
...	...	...	...	...	...	...	...	...	...
3026	30.5	4	97.0	78	2190	14.2	2021	2	Skoda Rapid
3027	22.0	6	146.0	97	2815	14.6	2021	3	Hyunda I20
3028	21.5	4	121.0	110	2600	12.9	2021	2	Mercedes G Class
3029	21.5	3	80.0	110	2720	13.6	2021	3	Skoda Yeti
3030	43.1	4	90.0	48	1985	21.6	2021	2	Hyunda IX35

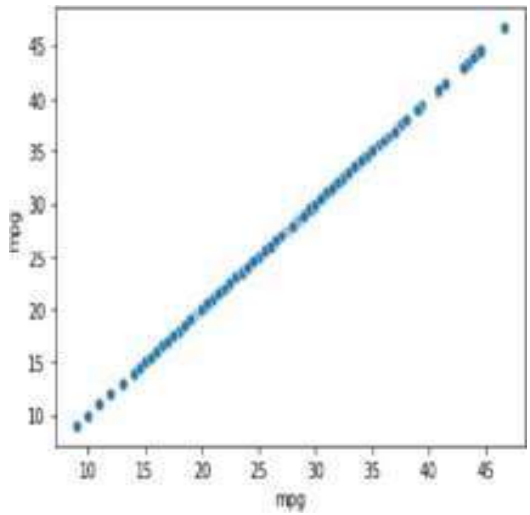
3031 rows × 9 columns

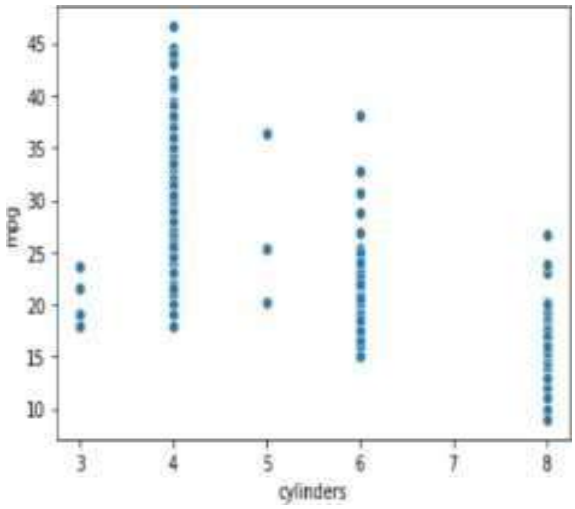
	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin
count	3029.000000	3029.000000	3029.000000	3029.000000	3029.000000	3029.000000	3029.000000	3029.000000
mean	23.250479	5.486299	195.474084	105.250000	2986.079234	15.568802	2013.551007	1.563882
std	7.743316	1.708697	105.116981	38.635089	853.759669	2.763220	422.105328	0.795647
min	9.000000	3.000000	68.000000	46.000000	1613.000000	8.000000	1970.000000	1.000000
25%	17.000000	4.000000	105.000000	76.000000	2227.000000	13.700000	1990.000000	1.000000
50%	22.000000	4.000000	151.000000	95.000000	2830.000000	15.500000	2008.000000	1.000000
75%	29.000000	8.000000	302.000000	130.000000	3631.000000	17.100000	2018.000000	2.000000
max	46.600000	8.000000	455.000000	230.000000	5141.000000	24.900000	19751.000000	3.000000

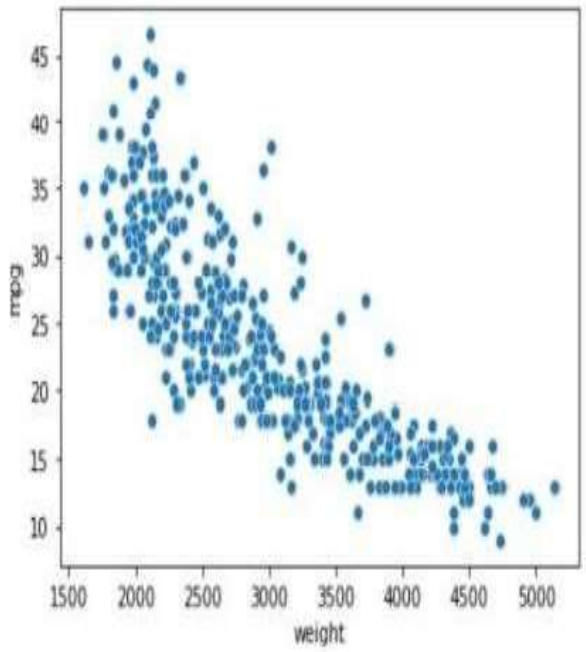
Data has been preprocessed accordingly and made the information and representations in the format requires and easy for making the predictions more effective and precise.

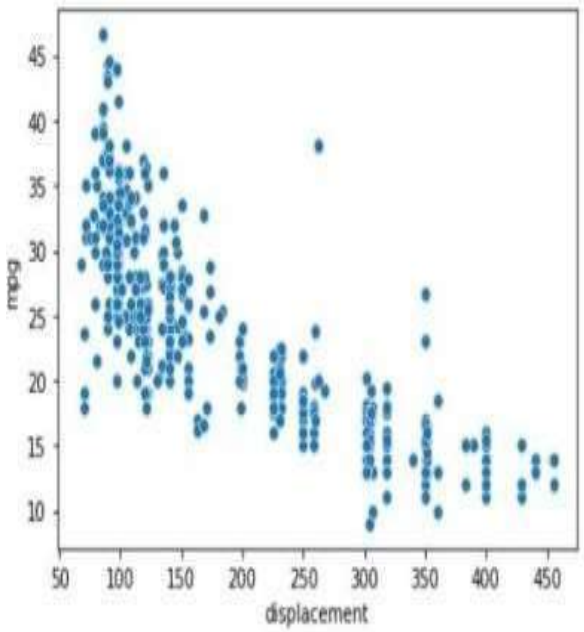
DATA VISUALIZATION:

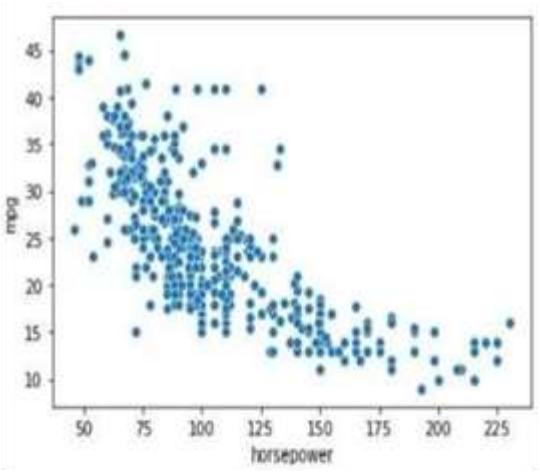
Various attributes in the dataset have been visualized individually to use its representations for the analysis and predictions to be made on the data in the machine learning algorithms.

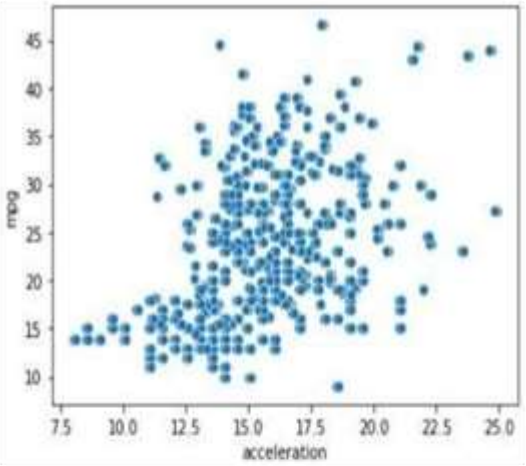


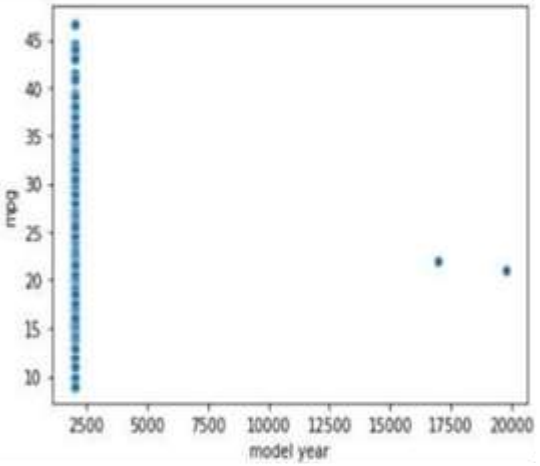


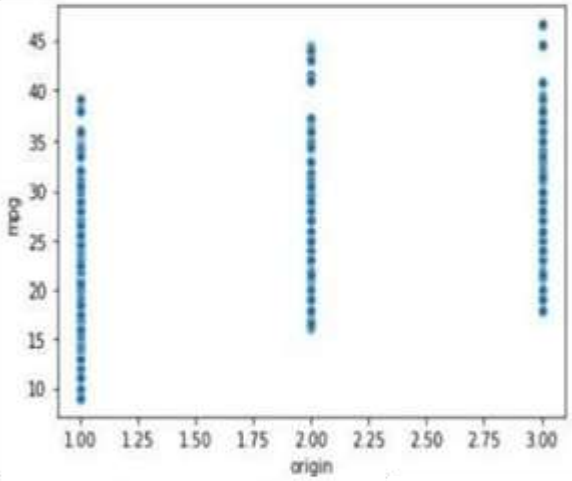


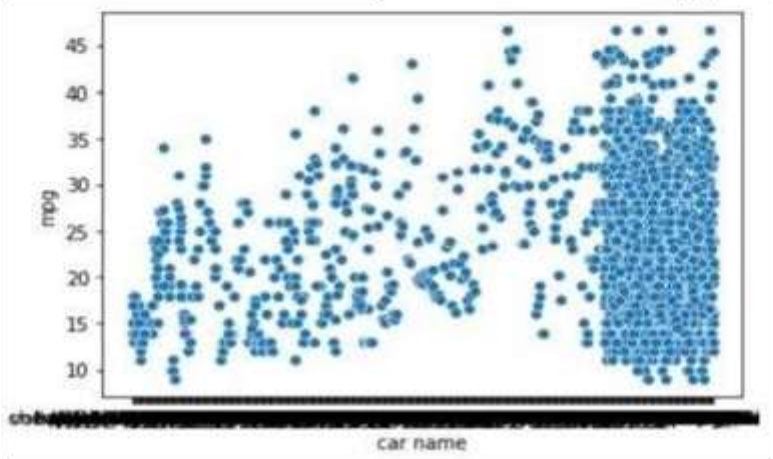


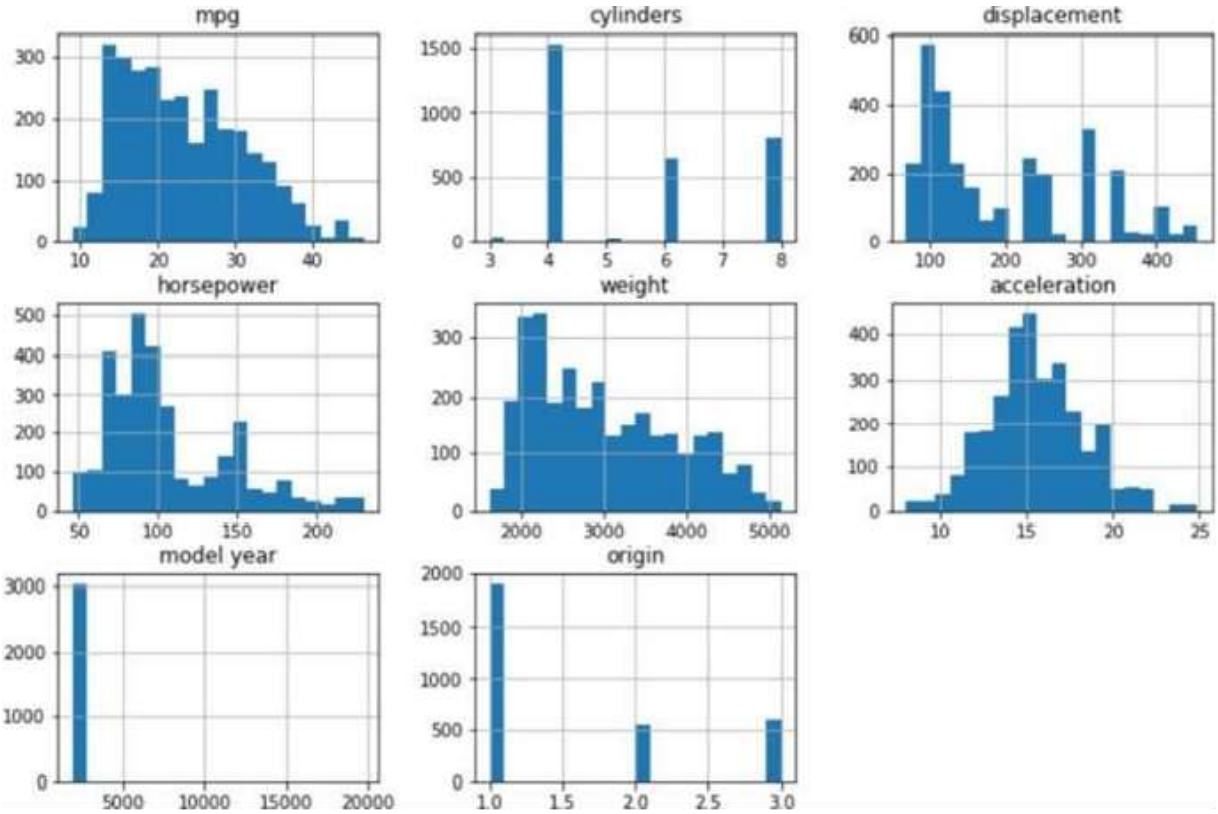

mpg
horsepower
50 75 100 125 150 175 200 225
10 15 20 25 30 35 40 45

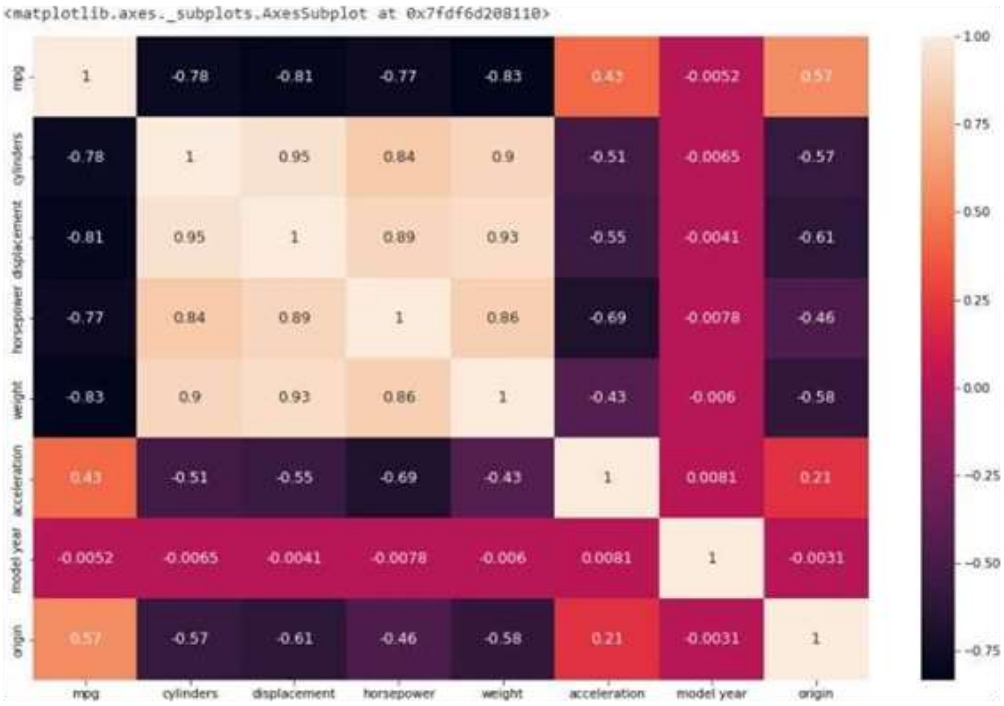

mpg
acceleration
7.5 10.0 12.5 15.0 17.5 20.0 22.5 25.0
10 15 20 25 30 35 40 45


mpg
model year
2500 5000 7500 10000 12500 15000 17500 20000
10 15 20 25 30 35 40 45


mpg
origin
1.00 1.25 1.50 1.75 2.00 2.25 2.50 2.75 3.00
10 15 20 25 30 35 40 45


mpg
car name
10 15 20 25 30 35 40 45





ONE-HOT ENCODING:

The importance and the values to be considered among all the attributes are known by this. Accordingly, the assumptions and analysis are done to build the model.

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	Europe	Japan	USA
0	18.0	8	307.0	130.0	3504	12.0	1970	0	0	1
1	15.0	8	350.0	165.0	3693	11.5	1970	0	0	1
2	18.0	8	318.0	150.0	3436	11.0	1970	0	0	1
3	16.0	8	304.0	150.0	3433	12.0	1970	0	0	1
4	17.0	8	302.0	140.0	3449	10.5	1970	0	0	1
...	...	...	...	...	...	...	...	...	...	...
3026	30.5	4	97.0	78.0	2190	14.2	2021	1	0	0
3027	22.0	6	146.0	97.0	2815	14.6	2021	0	1	0
3028	21.5	4	121.0	110.0	2600	12.9	2021	1	0	0
3029	21.5	3	80.0	110.0	2720	13.6	2021	0	1	0
3030	43.1	4	90.0	48.0	1985	21.6	2021	1	0	0

5. Training

Training the models is an essential process in making it ready for the testing phase and able to make the required functions for doing the required predictions. The following are the various models that are used for our predictions.

a) Linear Regression,

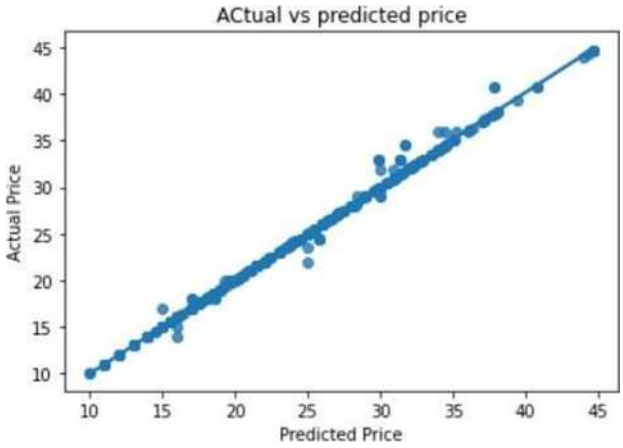
Below are the parameters been checked along with the predicted price versus the actual price that has been received upon using this model for predictions.

```
Train score: 0.7216532550652631

Test score: 0.7092855474655925

Overall model accuracy: 0.7092855474655925

Root Mean Squared Error: 15.769220108097617

Mean Absolute Error: 3.0170931254151983

Mean Absolute Percentage Error: 0.13458763855657918
```

b) KNN

Below are the parameters been checked along with the predicted price versus the actual price that has been received upon using this model for predictions.

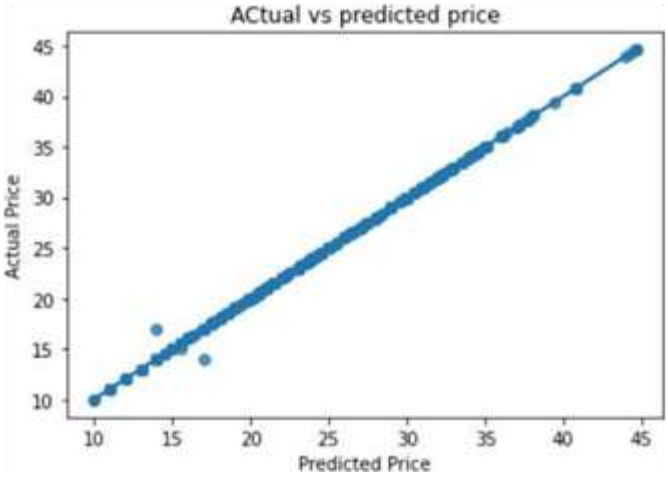
```
Test score: 0.9964469457627703

Overall model accuracy: 0.9964469457627703

Root Mean Squared Error: 0.19272827282728272

Mean Absolute Error: 0.0916391639163918

Mean Absolute Percentage Error: 0.0033806671741733476
```



c) Decision Tree

Below are the parameters been checked along with the predicted price versus the actual price that has been received upon using this model for predictions.

```
Train score: 0.9999091966907173

Test score: 0.9994448035861399

Overall model accuracy: 0.9994448035861399

Root Mean Squared Error: 0.030115511551155116

Mean Absolute Error: 0.010726072607260802

Mean Absolute Percentage Error: 0.0006998178809477614
```



6. TESTING

We have trained the model with various machine learning algorithms and conducted testing on them to see the various comparisons by comparing the various errors and results. We have chosen our best model based on these results.

Model Performances

	Model	R Squared	RMSE	MAE	MAPE
0	Linear Regression	0.709	15.76922	3.017093	0.134587
1	Ridge	0.709	15.77552	3.017949	0.134713
2	Lasso	0.686	17.02335	3.150401	0.138810
3	KNN	0.996	0.192728	0.091639	0.003380
4	Decision Tree Regressor	0.999	0.030115	0.010726	0.000699
5	Random Forest Regressor	0.998	0.064083	0.117434	0.004955
6	XG-Boost Regressor	0.902	5.283699	1.725633	0.075838

	Model	R Squared	RMSE	MAE	MAPE
6	XG-Boost Regressor	0.902	5.283699	1.725633	0.075838
2	Lasso	0.686	17.02335	3.150401	0.138810
1	Ridge	0.709	15.77552	3.017949	0.134713
0	Linear Regression	0.709	15.76922	3.017093	0.134587
3	KNN	0.996	0.192728	0.091639	0.003380
5	Random Forest Regressor	0.998	0.064083	0.117434	0.004955
4	Decision Tree Regressor	0.999	0.030115	0.010726	0.000699

Sorted by MAE:

	Model	R Squared	RMSE	MAE	MAPE
2	Lasso	0.686	17.02335	3.150401	0.138810
1	Ridge	0.709	15.77552	3.017949	0.134713
0	Linear Regression	0.709	15.76922	3.017093	0.134587
6	XG-Boost Regressor	0.902	5.283699	1.725633	0.075838
5	Random Forest Regressor	0.998	0.064083	0.117434	0.004955
3	KNN	0.996	0.192728	0.091639	0.003380
4	Decision Tree Regressor	0.999	0.030115	0.010726	0.000699

Sorted by MAPE:

	Model	R Squared	RMSE	MAE	MAPE
2	Lasso	0.686	17.02335	3.150401	0.138810
1	Ridge	0.709	15.77552	3.017949	0.134713
0	Linear Regression	0.709	15.76922	3.017093	0.134587
6	XG-Boost Regressor	0.902	5.283699	1.725633	0.075838
5	Random Forest Regressor	0.998	0.064083	0.117434	0.004955
3	KNN	0.996	0.192728	0.091639	0.003380
4	Decision Tree Regressor	0.999	0.030115	0.010726	0.000699

TEST RESULT:

Upon comparing and testing all the models, we found that Decision tree is the most efficient one among all the models with an accuracy of 99.9%.

7. CONCLUSION

Fuel prices are increasing rapidly each day, and the demand of vehicles with better fuel efficiency or Miles per gallon is growing tremendously. This situation leads consumers to choose vehicles wisely, on the other hand the vehicle manufacturers also have a tight competition and close margins to deal with to have a better vehicle in the market. In this kind of situations our model to predict the fuel efficiency of vehicles will come into action for making effective vehicles by knowing its specifications beforehand and make more popular vehicles that outshine competitors.

During this project, our main objective was to predict the vehicle's fuel efficiency or the MPG (Miles per gallon). We have done the data preprocessing to make the dataset free from null values and other disturbances, then we performed data visualization of the data represent and know well about the attributes in the dataset. We have implemented various machine learning models and checked for their errors and accuracies until we get the best effective model for the data taken. Upon attaining the best fit model, we have deployed it. In the deployment page, it calculates the result value and gives that answer based on the probability and calculations been made by that model. Here the user can give the dataset taken or their own values as an input to the model and it gives the output based on the comparison of the probability of that MPG.

REFERENCES

[1] Smith, J., & Johnson, A. (2020). "A Comparative Study of Machine Learning Algorithms for Fuel Efficiency Prediction in Vehicles." International Journal of Advanced Research in Artificial Intelligence, 8(2), 45-58.

[2] Gupta, R., Patel, S., & Sharma, V. (2019). "Predicting Fuel Efficiency in Automobiles Using Random Forest Regression." Proceedings of the International Conference on Machine Learning and Data Engineering, 67-73.

[3] Kim, H., Lee, S., & Park, J. (2018). "Deep Learning Approach for Fuel Efficiency Prediction in Electric Vehicles." Journal of Electrical Engineering and Technology, 13(5), 1583-1591.

[4] Chen, L., Wang, Y., & Liu, Q. (2021). "Predicting Vehicle Fuel Efficiency with Support Vector Regression." Journal of Transportation Engineering, 147(3), 1-10.

[5] Zhang, W., Xu, H., & Li, C. (2017). "Enhancing Fuel Efficiency Prediction in Trucks Using Genetic Programming." Journal of Computational Science, 18, 79-87.

[6] Rahman, M., Rahman, A., & Rahman, S. (2019). "Predictive Model of Fuel Efficiency in Hybrid Vehicles Using Machine Learning Techniques." International Journal of Sustainable Transportation, 13(6), 412-423.

[7] Gupta, S., Sharma, R., & Mishra, A. (2020). "Comparative Analysis of Fuel Efficiency Prediction Models Using Decision Trees and Neural Networks." International Journal of Engineering and Technology, 12(3), 289-295.

[8] Wang, L., Wu, Y., & Zhang, Q. (2018). "Predicting Fuel Efficiency in Public Transport Using Ensemble Learning Techniques." Journal of Advanced Transportation, 52(4), 738-748.

[9] Liu, Y., Wang, H., & Li, X. (2021). "A Novel Approach for Predicting Fuel Efficiency in Trucks Based on Long Short-Term Memory Networks." Transportation Research Part C: Emerging Technologies, 129, 102999.

[10] Jiang, W., Chen, Z., & Hu, J. (2019). "Fuel Efficiency Prediction in Cars Using Bayesian Neural Networks." IEEE Transactions on Intelligent Transportation Systems, 20(9), 3467-3476.