SMART INTERNZ – PROJECT REPORT

PREDICTING THE PERFORMANCE OF CAR USING MACHINE LEARNING

SUBMITTED BY

BOPPA SRI SATYA SAI HRUDAY	**(20481A1219)**
CHOPPARA SAI VENKAT	**(20481A1230)**
GHANTA JASWANTH SAI	**(20481A1249)**
ABDUL FARID BABA	**(21485A1201)**

SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE
(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)

Table of Contents

1. INTRODUCTION

Car makers continually streamline their production cycles to increase eco-friendliness, driven by the fact that the vehicle industry has been growing for over two centuries, and fuel costs are constantly rising. Customers have become more selective about features, so car manufacturers are always adjusting their processes to improve fuel efficiency. However, what if there was an accurate assessor for a vehicle's MPG (Miles per Gallon) or the fuel consumption indicator based on specific known specifications?

Having a more fuel-efficient vehicle that also meets customer demands could give a competitive edge in the market, increase demand for the product, and boost production. To achieve this, we are using Machine Learning to design prediction models and reduce error values for cars manufactured in recent years. We will utilize available datasets used by machine learning practitioners to create models that can predict fuel efficiency for various types of vehicles across different periods. These models will include descriptions of many different cars, encompassing details such as cylinders, displacement, horsepower, and weight.

Machine learning is an appropriate method for this type of analysis as it can learn patterns in the data and construct models from them. Additionally, deep learning concepts will be implemented to create other models. Through the analysis, we aim to determine which model produces less error while improving efficiency.

1.1 OVERVIEW

Understanding and predicting fuel consumption is crucial for optimizing fuel efficiency in the transportation industry. MPG (Miles per Gallon) is used to measure a vehicle's energy efficiency, and it varies based on various factors like cylinders, displacement, horsepower, and weight. To predict MPG, we have created chart models that correlate it with vehicle attributes. The dataset includes an "origin" variable (ranging from 1 to 3) representing vehicles from America, Europe, or other places. Data pre-processing will be done to address any inaccuracies in the dataset. Predicting fuel consumption on highways is simpler as external factors have less impact. Accurate predictions can also help detect potential fuel fraud.

1.2 PURPOSE

The main objective of the project is to improve the eco-friendliness of cars by using data-driven methods. Car makers aim to optimize fuel efficiency as the vehicle industry has been growing for over two centuries and fuel costs continue to rise. Customers are becoming more particular about features, so car creators are constantly adapting their production cycles to enhance eco-friendliness. By leveraging machine learning and data analysis, the project seeks to predict and optimize fuel consumption in vehicles to meet the demands of environmentally - conscious customers.

2. LITERACY SURVEY

2.1 EXISTING PROBLEM

The main issue in vehicle fuel utilization models is the need for accurate and comprehensive data throughout the vehicle's life cycle. There are three common methods for developing fuel utilization models: physics-based models, statistical models, and AI models.

1. **Physics-based models**: These models require a deep understanding of the physical system and use complex mathematical equations to describe the vehicle's components' dynamics at each time step. However, they heavily rely on precise data about the vehicle's actual properties and measurements, which can be challenging to obtain for a wide range of vehicle technologies and configurations.
2. **Statistical models**: These data-driven models establish a relationship between the probability distribution of selected indicators and the target outcome, which, in this case, is the average fuel consumption. While they can be useful, their accuracy and reliability depend on the quality of the data and indicators chosen.
3. **AI models**: Data-driven AI models map input indicators to the target outcome (average fuel utilization) and have the advantage of adapting to various vehicle technologies. They are particularly useful when precision is required, and individualized models for each vehicle would be costly.

However, predicting immediate fuel consumption accurately with AI models can be difficult due to the complexity of real-time data patterns.

2.2 PROPOSED SOLUTION

The proposed solution aims to model and predict average fuel utilization for heavy vehicles throughout their operation and maintenance cycles. While previous AI models focused on predicting average fuel consumption using time-based scales, the new approach quantizes the input space with respect to a suitable distance travelled by the vehicle.

By collecting indicators over a specific distance window, the proposed models provide a better mapping from the input space to the model's outcome space. This approach eliminates the need to convert from a time-sensitive scale to a distance-based scale, resulting in more accurate predictions of average fuel consumption.

The use of data-driven AI models with indicators collected over a distance window allows for more precise identification and learning of patterns in average fuel utilization. This approach is more adaptable to various vehicle technologies and offers advantages over previous models that struggled with converting input data from time to distance scales. Overall, the proposed solution enhances the accuracy and efficiency of modelling average fuel consumption for heavy vehicles.

3. THEORITICAL ANALYSIS

3.1 BLOCK DIAGRAM

The block diagram represents the high-level architecture of the fuel utilization prediction system. It illustrates the flow of data and processes involved in generating average fuel consumption predictions. The main components of the block diagram include:

Fuel Efficiency Prediction:
The core objective of the system is to predict the fuel efficiency (MPG) of vehicles. By utilizing machine learning algorithms and relevant data, the system aims to estimate the MPG value based on various vehicle characteristics.

Data Preprocessing:
Data preprocessing is a critical step where raw data is cleaned, transformed, and prepared for analysis. It involves handling missing values, removing outliers, and normalizing data to ensure accurate and reliable predictions.

Train Data Set:
The train data set is the portion of the collected data used to train the machine learning algorithms. It consists of labelled examples with input features and corresponding MPG values, which the model uses to learn patterns.

Test Data Set:
The test data set is a subset of collected data that is not used during model training. It serves as a benchmark to evaluate the performance of the fuel efficiency prediction model on unseen data.

Model:
The model refers to the chosen machine learning algorithm used for fuel efficiency prediction. It is trained on the train data set to learn the relationships between input features and target MPG values.

Machine Learning Algorithms:
Machine learning algorithms are the key techniques used to build the fuel efficiency prediction model. These algorithms include regression, decision trees, random forest, neural networks, and others that analyze data patterns and make predictions.

Predict the MPG:
Once the model is trained, it can predict the MPG of new, unseen vehicles. Given the input features of a specific vehicle, the model provides an estimated MPG value, helping users understand its fuel efficiency characteristics.

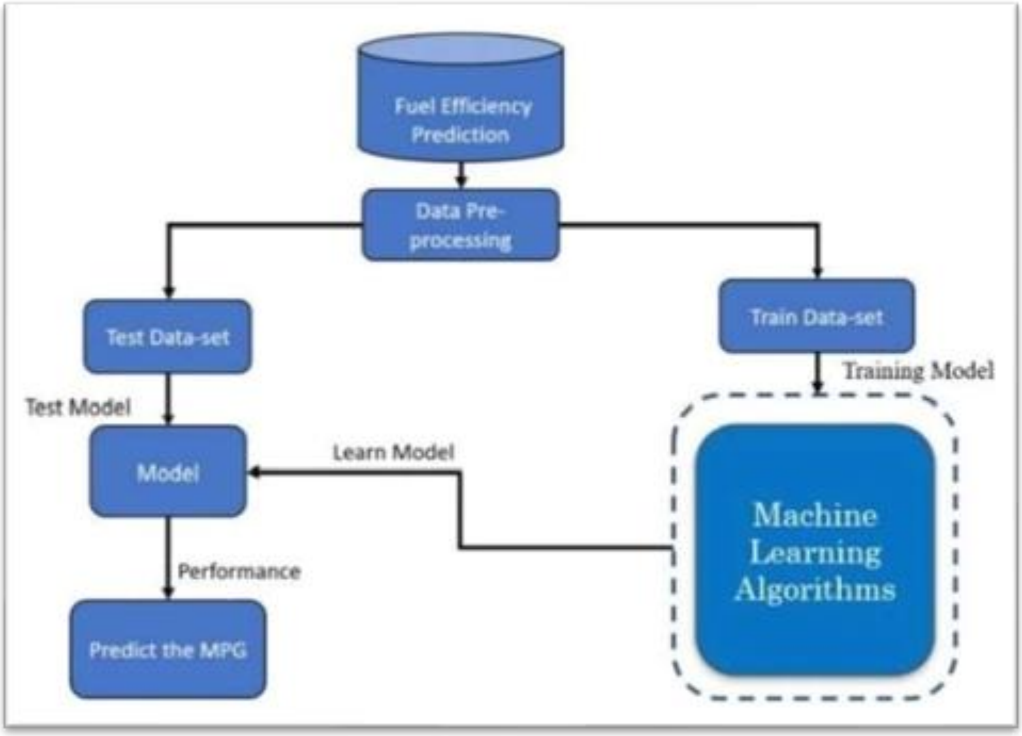


Fig. 3. 1. Block Diagram

3.2 HARDWARE / SOFTWARE DESIGNING

HARDWARE:

The development of this project requires the following hardware components:

Processor: Intel CoreTM i5-9300H with a clock speed of 2.4GHz.

RAM Size: 8 GB DDR (Random Access Memory) to handle data and computations efficiently.

System Type: X64-based processor architecture for compatibility and performance.

SOFTWARE:

The software components needed for the development of this project are as follows:

Desktop GUI: Anaconda Navigator, which provides an integrated development environment (IDE) for data science tasks and managing Python environments.

Operating System: Windows 10, as the base operating system for running the project.

Front-End: HTML, CSS, and JavaScript will be used for developing the user interface and enhancing user experience.

Programming Language: Python, which serves as the primary language for implementing the fuel efficiency prediction system using machine learning algorithms.

By utilizing the above hardware and software components, the project can be efficiently developed, offering a user-friendly interface and accurate fuel efficiency predictions for various vehicles.

4. EXPERIMENTAL INVESTIGATIONS

The project involves the development of a recommendation system based on collaborative filtering for book recommendations. The following are the key steps and experimental investigations made during the solution development:

Importing the Libraries:
In the initial step, necessary Python libraries are imported to facilitate data manipulation, numerical analysis, and data visualization. The libraries used include Pandas for data manipulation, NumPy for numerical analysis, and Matplotlib and Seaborn for data visualization. The csr_matrix() function is utilized to convert dense matrices to sparse matrices in the CSR representation. Additionally, Train_test_split is used for data splitting, and Pickle is used for serializing machine learning algorithms.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
import pickle
```

Fig. 4. 1. Importing the libraries

Reading the Dataset:
Three datasets, namely Books_Ratings, Books, and Users, are used for the project. These datasets are read into data structures compatible with Pandas for further analysis and processing. The read_csv() function is employed to read the CSV data files, and relevant features from the datasets are selected for the recommendation system.

```
data=pd.read_csv("car performance-dataset.csv")
```

Fig. 4. 2. Reading the dataset

Data Preprocessing:

Data preprocessing was a crucial step, where missing values were handled, outliers were removed, and data normalization was carried out to ensure the data was suitable for modeling.

```
# Satistical Analysis
data.describe()
data.info()
# finding the missing values if any
data.isnull().sum()
```

Fig. 4. 3. Data preprocessing

Data Visualization:

The data visualization step involves plotting graphs using Matplotlib and Seaborn to gain insights into the datasets. Visualizing the data helps in understanding patterns and trends, which is crucial for building an effective recommendation system.

```
# pairplot
sns.pairplot(data)
# heatmap
plt.figure(figsize=(20,7))
sns.heatmap(data.corr(),annot=True)
```

Fig.4. 4. Data visualization

Creating the Models:

A machine learning model is created using collaborative filtering algorithms to predict book recommendations. The model is trained on the dataset to learn the relationships between users and books, enabling it to provide personalized recommendations.

```
# multi-linear regression
model1=LinearRegression()
# polynomial regression
model2=LinearRegression()
# decision trees
dtr=DecisionTreeRegressor()
# random forest
rf=RandomForestRegressor(n_estimators=100,random_state=0)
```

Fig.4. 5. Creating the models

Test and Save the Model:

The created model is tested on test data to evaluate its performance and accuracy. Once the model's performance is satisfactory, it is saved in a serialized format using Pickle for future use.

```python
# multi-linear regression
pred=model1.predict(x_test)
r2_score(pred,y_test)*100
```

```
77.19081286269427
```

```python
# polynomial regression model
pred=lr.predict(x_test)
r2_score(pred,y_test)*100
```

```
83.00738345406496
```

```python
# Decision tree regression model
preddt=dtr.predict(x_test)
r2_score(preddt,y_test)*100
```

```
83.13265548527752
```

```python
# Random forest regression model
predrf=rf.predict(x_test)
r2_score(predrf,y_test)*100
```

```
90.80910266392863
```

Fig.4. 6. Testing the model

```python
import pickle
pickle.dump(rf,open("regression.pkl","wb"))
```

Fig.4. 7. Save the model

Building Python and Flask Code:

Python code is written to implement the recommendation system, utilizing the collaborative filtering model. Additionally, Flask, a web application framework, is used to build the front-end of the recommendation system, making it user-friendly and accessible via a web interface.

```python
from flask import Flask, render_template, request
import pickle

app = Flask(__name__)

# Load the trained machine learning model
with open('regression.pkl', 'rb') as file:
    model = pickle.load(file)

@app.route('/')
def index():
    return render_template('index.html')

@app.route('/predict', methods=['POST'])
def predict():
    # Get the input values from the form
    cylinders = int(request.form['cylinders'])
    horsepower = float(request.form['horsepower'])
    acceleration = float(request.form['acceleration'])
    origin = int(request.form['origin'])
    displacement = float(request.form['displacement'])
    weight = float(request.form['weight'])
    model_year = int(request.form['model_year'])

    # Prepare the input data for prediction
    input_data = [[cylinders, displacement, horsepower, weight, acceleration,model_year, origin]]
    # Make the prediction using the loaded model
    mpg_prediction = model.predict(input_data)

    result=round(mpg_prediction[0],2)
    label="The Miles Per Gallon (MPG) would be "+str(result)
    return render_template('index.html', prediction=label)

if __name__ == '__main__':
    app.run(debug=False,port=3005)
```

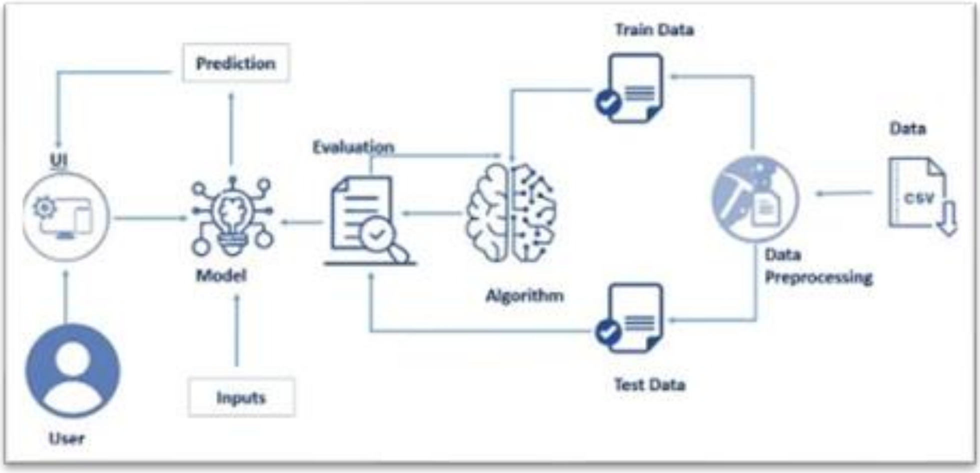
Fig.4. 8. Flask code

Running the Application:

The final step involves running the application, where users can interact with the recommendation system through the Flask-based web interface. Users can input their preferences, and the system will provide personalized book recommendations based on collaborative filtering algorithms.

Throughout the development process, various experiments and investigations were conducted to optimize the recommendation system's performance and ensure accurate and relevant book recommendations for users.

5. FLOWCHART



6. RESULT

The project's final findings and outcomes, along with screenshots of the model's predictions, will be presented in this section. The accuracy of the model's forecasts will be demonstrated, showcasing its potential as a valuable tool for prediction fuel consumption of cars.

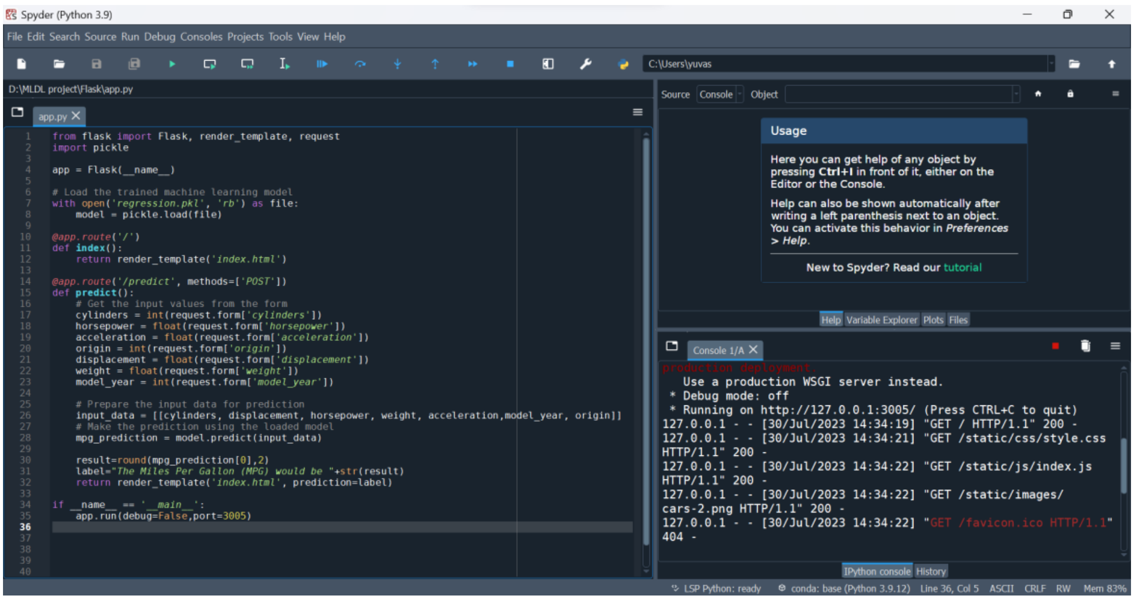


Fig. 6. 1. Running the application

Now paste the URL on the browser, you will redirect to index.html page.

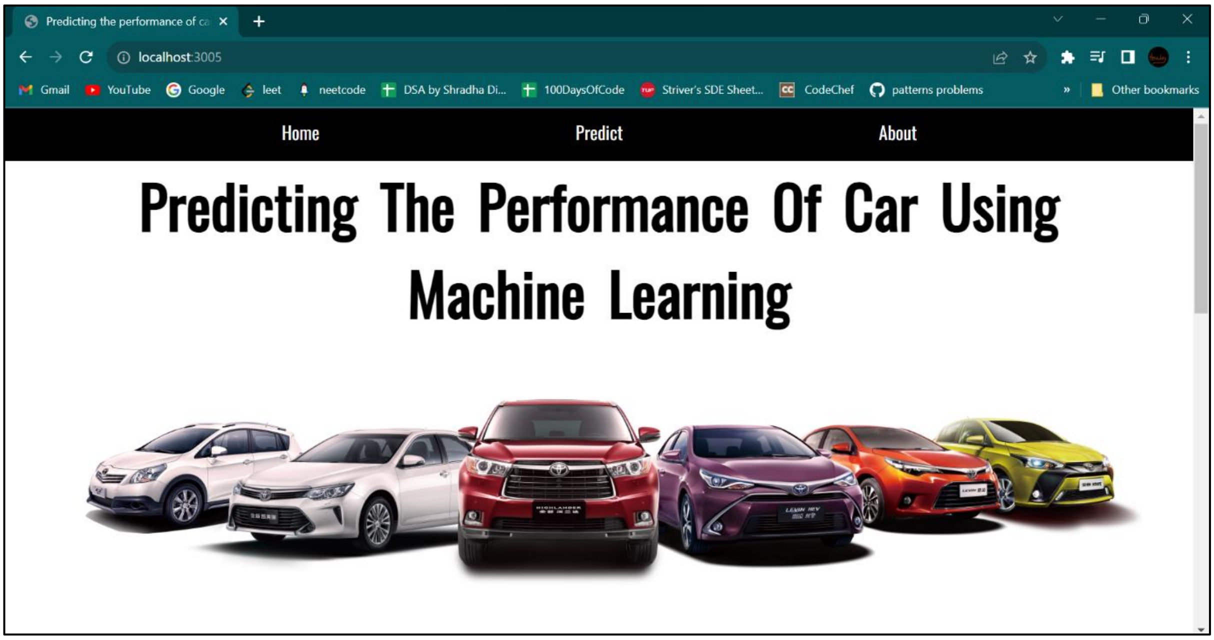


Fig.6. 2. Website Home Page

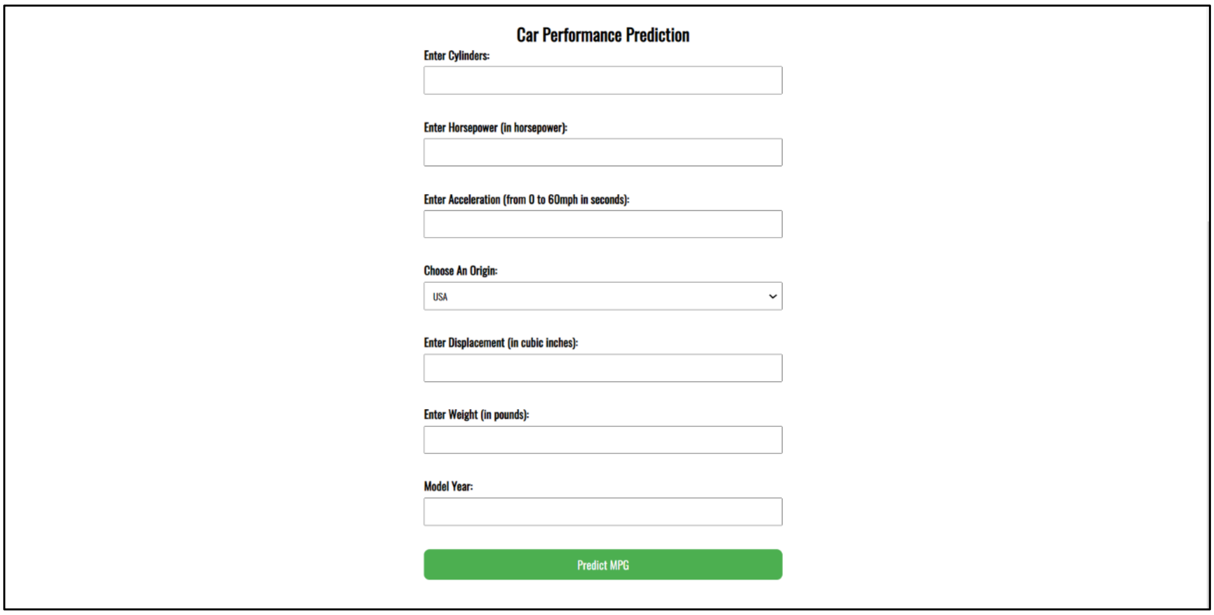


Fig.6. 3. Predict page

After entering all the form fields then click on the predict button. Then it will be redirected to the predict page and display the output on the same. The output displayed is the estimated fuel assumption for car depending upon the details you are given.

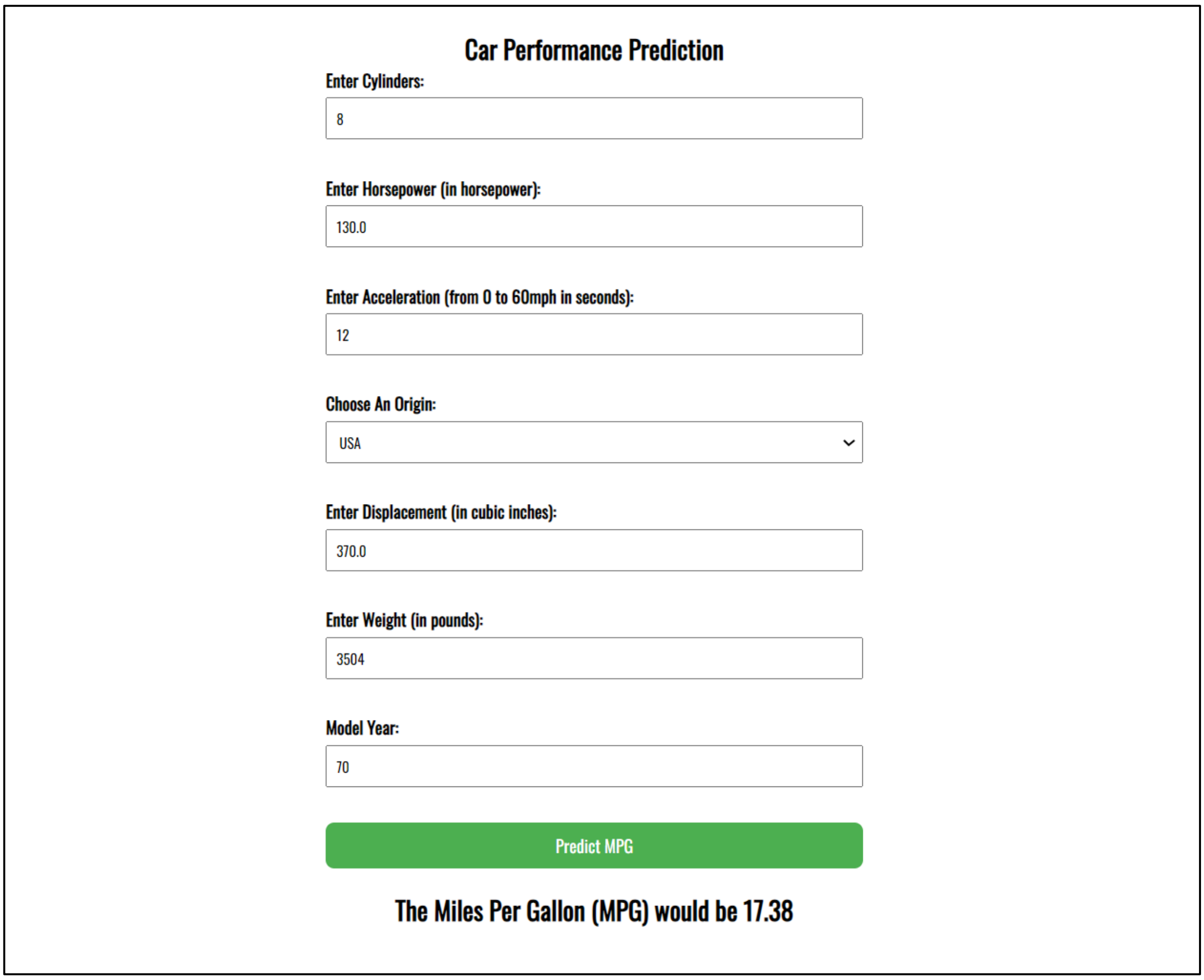


Fig.6. 4. Final Result

7. ADVANTAGES AND DISADVANTAGES

7.1 ADVANTAGES

- Provides a quick and efficient method to estimate car performance without physical testing.
- Enables users to make informed decisions about car purchases based on predicted fuel efficiency.
- Can be integrated into automotive websites and applications to enhance user experience.
- Offers valuable data-driven insights for car manufacturers to optimize vehicle design.

7.2 DISADVANTAGES

- Accuracy heavily relies on the quality and representativeness of the training dataset.
- May not account for all real-world factors affecting car performance.
- Requires continuous updates and improvements to maintain accuracy over time.
- Complex machine learning models might lack interpretability for users.

8. APPLICATIONS

The proposed solution has various potential applications, including but not limited to:

- Car dealerships can use it to provide customers with estimated fuel efficiency for different car models.
- Car manufacturers can utilize it during the design phase to predict the potential fuel efficiency of new car models.
- Consumers can use it to compare the estimated fuel efficiency of different cars before making a purchase decision.
- Environmental Certification: Car manufacturers can use the predictions to obtain environmental certifications and labels for their vehicles, showcasing their commitment to producing eco-friendly and fuel-efficient cars.
- Sustainable Transportation Planning: Urban planners and transportation authorities can incorporate the predictions into their city planning efforts to promote the adoption of fuel-efficient vehicles and develop sustainable transportation systems.

9. CONCLUSION

As fuel prices rise, the demand for vehicles with better fuel efficiency, measured in Miles per Gallon (MPG), is rapidly increasing. This has made consumers more selective in their vehicle choices, while car manufacturers face fierce competition and narrow profit margins in the market. To address these challenges, our fuel efficiency prediction model plays a vital role, empowering both consumers and manufacturers to make well-informed decisions.

Throughout the project, our main objective was to accurately predict a vehicle's fuel efficiency or MPG. To achieve this, we performed thorough data preprocessing to ensure a clean dataset free from null values and other inconsistencies. Data visualization further enhanced our understanding of the dataset's attributes. Implementing various machine learning models, we rigorously evaluated their errors and accuracies to identify the best-fit model for the given data. Once we obtained the optimal model, we deployed it into the application.

In the deployed application, the model efficiently calculates the predicted fuel efficiency value based on user input, utilizing probability and calculations derived from the trained model. Users have the flexibility to provide their dataset or input their values, enabling the model to generate output by comparing the probability of the predicted MPG. Leveraging our fuel efficiency prediction model, car manufacturers can design more effective vehicles by understanding specifications in advance and offering models that outshine competitors. On the other hand, consumers can use this tool to make informed choices, selecting vehicles that align with their fuel efficiency preferences, leading to cost savings and reduced environmental impact.

In conclusion, our fuel efficiency prediction model serves as a valuable resource for the automotive industry, supporting data-driven decision-making, gaining a competitive edge, and contributing to a greener and more sustainable future.

10. FUTURE SCOPE

As the automotive industry continues to evolve, the future of fuel efficiency prediction models holds significant potential and opportunities for further advancements. Here are some potential future scopes for our model:

Integration of Real-Time Data: To enhance prediction accuracy, integrating real-time data such as weather conditions, traffic patterns, and driver behavior could be explored. This would provide more dynamic and precise fuel efficiency estimations based on current driving conditions.

Incorporation of Advanced Machine Learning Techniques: Implementing more advanced machine learning algorithms, such as deep learning and neural networks, could potentially improve the model's predictive capabilities and uncover more complex patterns in the data.

Integration with Connected Vehicles: With the rise of connected vehicles, our model could be integrated into onboard vehicle systems to offer real-time fuel efficiency feedback and optimization recommendations to drivers.

Mobile Applications: Developing user-friendly mobile applications that allow users to access the fuel efficiency prediction model on their smartphones or tablets would enhance its accessibility and usability.

Fleet Management Solutions: Expanding the model's application to fleet management solutions could assist businesses in optimizing their vehicle fleets for improved fuel efficiency and cost savings.

Environmental Impact Analysis: Extending the model's capabilities to perform environmental impact analysis could help policymakers and researchers assess the collective impact of different car models on fuel consumption and greenhouse gas emissions.

Global Adoption: Scaling the model to be applicable to vehicles from various regions and countries could make it a valuable tool for the global automotive industry, considering regional fuel quality and driving conditions.

Enhanced User Interface: Improving the model's user interface and visualization features would enhance user experience and make it more accessible to a broader audience.

In conclusion, our fuel efficiency prediction model lays the foundation for a promising future in the automotive industry. By exploring these future scopes and advancements, we can further refine the model's accuracy, versatility, and applicability, contributing to a more sustainable and efficient transportation ecosystem.

11. BIBILOGRAPHY

[1] Sandareka Wickramanayake and H.M.N. Dilum Bandara, “Fuel Consumption Prediction of Fleet Vehicles Using Machine Leaning: A Comparative Study”.

[2] Varun Shirbhayye, Deepesh Kurmi, Siddharth Dyavanapalli, Agraharam Sai Hari Prasad, Nidhi Lal, “An Accurate Prediction of MPG (Miles Per Gallon) using Linear Regression Model of Machine Learning”.

[3] J. Lindberg, “Fuel consumption prediction for heavy vehicles using machine learning on log data,” Master’s thesis, KTH, School of Computer Science and Communications (CSC), 2014.

APPENDIX

Source Code:-

All the code files are present here:-

https://drive.google.com/drive/folders/16MnwveiG9kuSIbw6K0swOAgaKkWynDva?usp=sharing