PROJECT REPORT

CAR PERFORMANCE PREDICTION USING IBM WATSON

Submitted by :
Rishikesh Suresh Kumar - 20BCE0462
Shikhar Kumar - 20BBS0215
Biggyat Kumar Pandey - 20BCE2763
Yash Pokerna - 20BBS0212

TABLE OF CONTENTS

INTRODUCTION

1.1 Overview

Car performance prediction using IBM Watson is a process that utilizes the capabilities of IBM Watson, to predict the performance of cars. Car performance prediction using IBM Watson offers a data-driven approach to understand and optimize the performance characteristics of vehicles, enhancing the driving experience and advancing the automotive industry.

1.2 Purpose

The purpose of car performance prediction using IBM Watson is to leverage advanced analytics and machine learning techniques to forecast the performance of cars. It aims to provide accurate and reliable predictions about various aspects of car performance, including acceleration, top speed, fuel efficiency, and handling. It helps car manufacturers, automotive engineers, and enthusiasts make informed decisions during the design, development, and tuning processes.

LITERATURE SURVEY

2.1 Existing approaches

Author & year	Proposed Approach	Dataset Used	Observations	Limitations
The Concept to Measure the Overall Car Performance Author: Jarut Kunanoppadol Year: Received: 25 January 2012 ; Accepted: 11 April 2012	first, we explain the information of a dynamometer; second, we present the theoretical car performance calculation method; third, the simulation results are shown; fourth, we present the conceptual implementation framework; and finally, we concludes with the discussion	Engine torque data at every operating speed are informed by the engine performance curve. But if we do not have We can calculate the output torque and power from the engine, and then simulate the engine curve, calculated simulation is needed. Which can be done with details of car speci-fi cation; maximum power, maximum torque, and engine speed at these points.	The overall car performance depends on various operating factors, such as the engine performance, transmission design, suspension optimization, car dimension and shape design, aerodynamic, friction reduction technology,and driver skill. An on-road experiment is necessary for a designer, driver, tuner, developer, and researcher to investigate the final result.Overall, car accelerate performance speeds up a car within the considering time.	The concept to measure the overall car performance from acceleration capability is possible and convenient because we can collect digital input signals from an exist-ing electronic control unit and transfer it to an additional processor for analyzing and displaying the final result in every mobile display, such as laptop, tablet, and smartphone. No doubt the concept is cost effective and easy for installation but the results from this research won't be precise, as so

				many other factors have been ignored for the research. Also the chances of acquiring the tampered data are high in this case.
Dynamic measuring of performance parameters for vehicles engines Authors: Martin-Pexa, Miroslav Müller, Sergej Hloch. Year: Revised 11 July 2017, Accepted 13 July 2017	Paper deals with dynamic measurement of performance parameters for selected vehicles engines using incremental sensors attached to the rest bed roller.The moment of the inertia is the result of the measurement, which can also be used in the measurement of performance parameters. The external speed characteristics of the engine primarily resulting from the dynamic measurement is the absolute result of the measurement. It can be concluded on the technical conditions of the motor vehicle based on the evaluation of external speed characteristics of the engine. It is possible to maintain a consistently good technical state of the motor engine with appropriate maintenance and contribute to the ecology and the economy of the vehicle operation.	There are several ways to obtain the moment of inertia of the engine: -Obtaining information from the manufacturer -Calculation -With makeweight -New engine -Dynamometer -Average	This method of accuracy detection is based on the measurement of a larger number of vehicles which are guaranteed by the producer that engine performance parameters reach certain values. It is a vehicle with mileage kilometers - maximally 30–40,000 km. Thus, the observed initial accuracy is around 5.5%. However, it is necessary to measure the number of larger vehicles. The moment of inertia is one of the necessary input	The measurement procedure can be applied to any passenger road vehicle. Restrictions are only related to the fact that the vehicle cannot be measured with a four-wheel drive with automatic transmission and the vehicle whose speed would have exceeded 140 km/h at the selected gear. There is no limitation of the roller test bed due to the dynamic principle of the measurement (the energy is dissipated by its own inertia of the running engine)

Mileage efficiency of cars Author: Neelesh A. Patankar , Jane Lin , Tanveer N. Patankar Year: Received 9 November 2020, Revised 9 June 2021, Accepted 5 August 2021	The goal of this paper is to propose a mileage efficiency metric to enable direct comparison of energy efficiency and greenhouse gas emission of cars for various vehicle mass, fuel type, vehicle make, vehicle technology, road infrastructure, and drive schedule. The concept of IC is then introduced, followed by the definition of the new mileage efficiency metric. The mileage efficiency metric is extended to an analytical CO2 emission model of a RWC and an IC is discussed. The potential applications and policy implications of the mileage efficiency metric and the analytical CO2 emission model are demonstrated in the remainder of the paper. Specifically, describes the analyses included in the paper and the data used for them. Then presents the results. Further policy implications of the findings are discussed. And lastly, conclusions are presented.	data including the actual mileage, emission, and road load data from current fleet are obtained from U.S. EPA's Test Car List data files (U.S. EPA, 2021c); historical fuel economy data by manufacturer are obtained from U.S. EPA's report on light-duty vehicle technology, CO_2 emissions, and fuel economy trends (U.S. EPA, 2021d).	Mileage efficiency is defined as the ratio of actual car fuel economy to that of an ideal brake-loss car (IBC), which is defined as the one without internal or external losses; or an ideal powertrain car (IPC), which is defined as the one without internal losses. As such, mileage efficiency is able to attribute efficiency loss of a Real World Car (RWC) to internal and/or external losses. It provides an alternative measure of energy efficiency of vehicles in terms of the energy efficiency potential that can be achieved with technology advancement, engineering design, infrastructure improvement, and government policies and incentives.	Building on the mileage efficiency metric, the paper further derived a fundamental framework that is capable of quantifying the effects of various important factors on vehicle energy consumption (and GHG emissions). Such a framework can be used to inform government policy making for vehicle technology, alternative fuel choice, and infrastructure investments. Future investigation will focus on how to devise effective policy to encourage consumer choice of alternative fuel vehicles as well as improve road infrastructure for more fuel efficient driving.

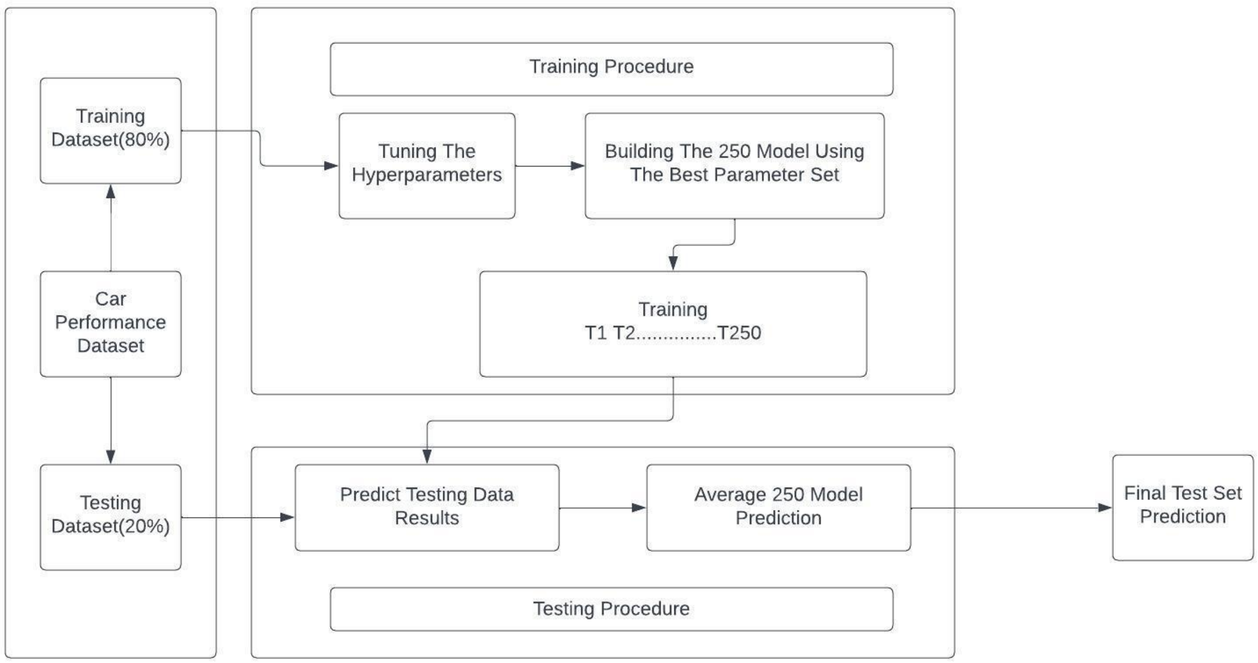
2.2 Proposed solution

The proposed solution focuses on developing a reliable car performance prediction model. Firstly, we preprocess the dataset to ensure its quality and suitability for analysis. This includes handling missing data, eliminating redundant features, and normalizing numerical variables. Next, we aim to identify and handle any anomalies or outliers present in the dataset. To predict car performance accurately, we utilize multiple regression models, specifically the Random Forest Regressor, K-Neighbors Regressor, Support Vector Regressor, and Multiple Linear Regressor. Each of these models has its unique characteristics and strengths, making them suitable for different scenarios. After training each regression model using appropriate training and validation techniques, we evaluate their performance using the Root Mean Squared Error (RMSE) metric. RMSE measures the average distance between the predicted values and the actual values. A lower RMSE indicates a more accurate prediction. By comparing the RMSE values of the different models, we can select the one that performs the best in terms of predicting car performance accurately.

THEORETICAL ANALYSIS

3.1 Block diagram

Diagrammatic overview of the project.



3.2 Hardware / Software designing

Hardware and software requirements of the project

Hardware Specifications:
The basic components like

- Processor: A modern processor with multiple cores for efficient data pre processing and other capabilities.
- Memory (RAM): Sufficient RAM to handle data processing and model training and testing.
- Wifi connections: for supporting the support of the jupyter notebook to run the required packages algorithms and other works regarding flask and the web based UI for the user to input the data to check the predictions of the model

Software Specifications:

- Programming: Knowledge of programming language Python, which is commonly used for data analysis and machine learning tasks, and performing various machine learning algorithms and use of tools like **Anaconda, spyder, jupyter notebook** would be essential also the use of **flask**
- Data Analysis Libraries: Python libraries such as pandas and NumPy are utilized for data cleaning, exploration,and manipulation.
- Machine Learning Libraries: Libraries like scikit-learn or label encoder train test split the accuracy score as a part of the evaluation metrics used for developing and training machine learning models.
- CSV Data Import: The project would involve importing the CSV file containing the car performance data. Python libraries like pandas can be used to read and manipulate the data.
- Data Visualization: Libraries like Matplotlib or Seaborn can be used for visualizing the data and generating insights.

EXPERIMENTAL INVESTIGATIONS

The experimental investigations in a car performance prediction project could involve various aspects depending on the specific goals and objectives of the project. Here are some possible experimental investigations that could be conducted:

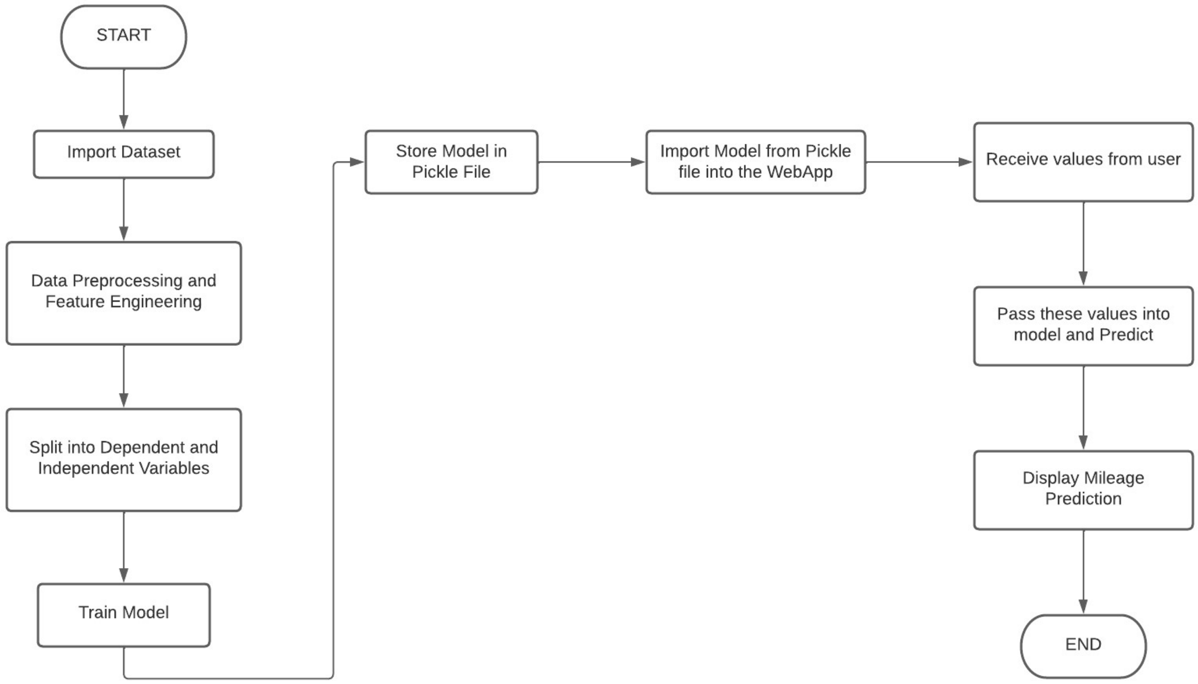
- Feature Selection: Experimentally determine the most influential features or variables like the independent and the dependent values of the given dataset for the train and test split that have a significant impact on car performance. It involves analyzing the correlation between different features and the performance metrics to identify the most relevant ones.

- Model Evaluation: different machine learning models such as linear regression, decision trees, random forests to determine which model performs best for predicting car performance. It involves training and evaluating multiple models and their combinations using appropriate evaluation metrics such as mean squared error,r2score for the accuracy test R-squared.

- Data Preprocessing Techniques: Experiment with various data preprocessing techniques such as normalization,regressors,RMSE, feature scaling, or handling missing values to determine the impact on model performance.

- Hyperparameter Tuning: Experiment involving different hyperparameter values for the selected machine learning models to optimize their performance. This would involve conducting grid search or random search to find the best combination of hyperparameters that yield the most accurate predictions.

FLOWCHART



START
Import Dataset
Data Preprocessing and Feature Engineering
Split into Dependent and Independent Variables
Train Model
Store Model in Pickle File
Import Model from Pickle file into the WebApp
Receive values from user
Pass these values into model and Predict
Display Mileage Prediction
END

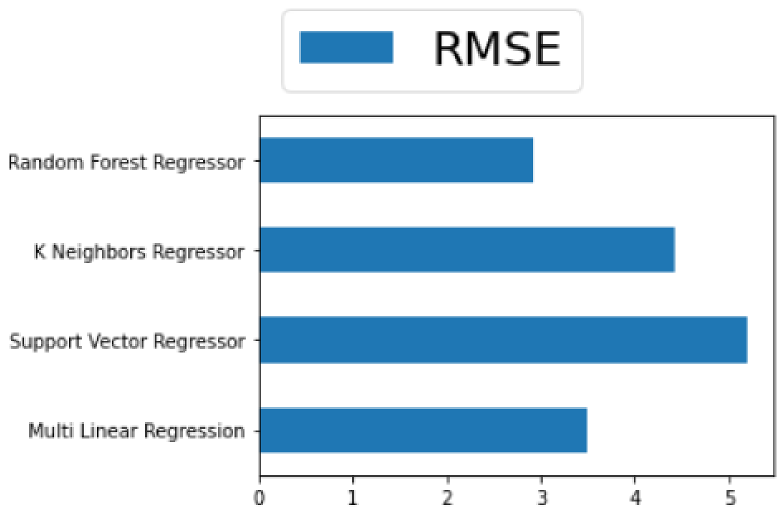
RESULT

In [47]:
```
df_model = pd.DataFrame(index=models.keys(), columns=['RMSE'])
df_model['RMSE'] = rmse.values()
df_model
```

Out[47]:

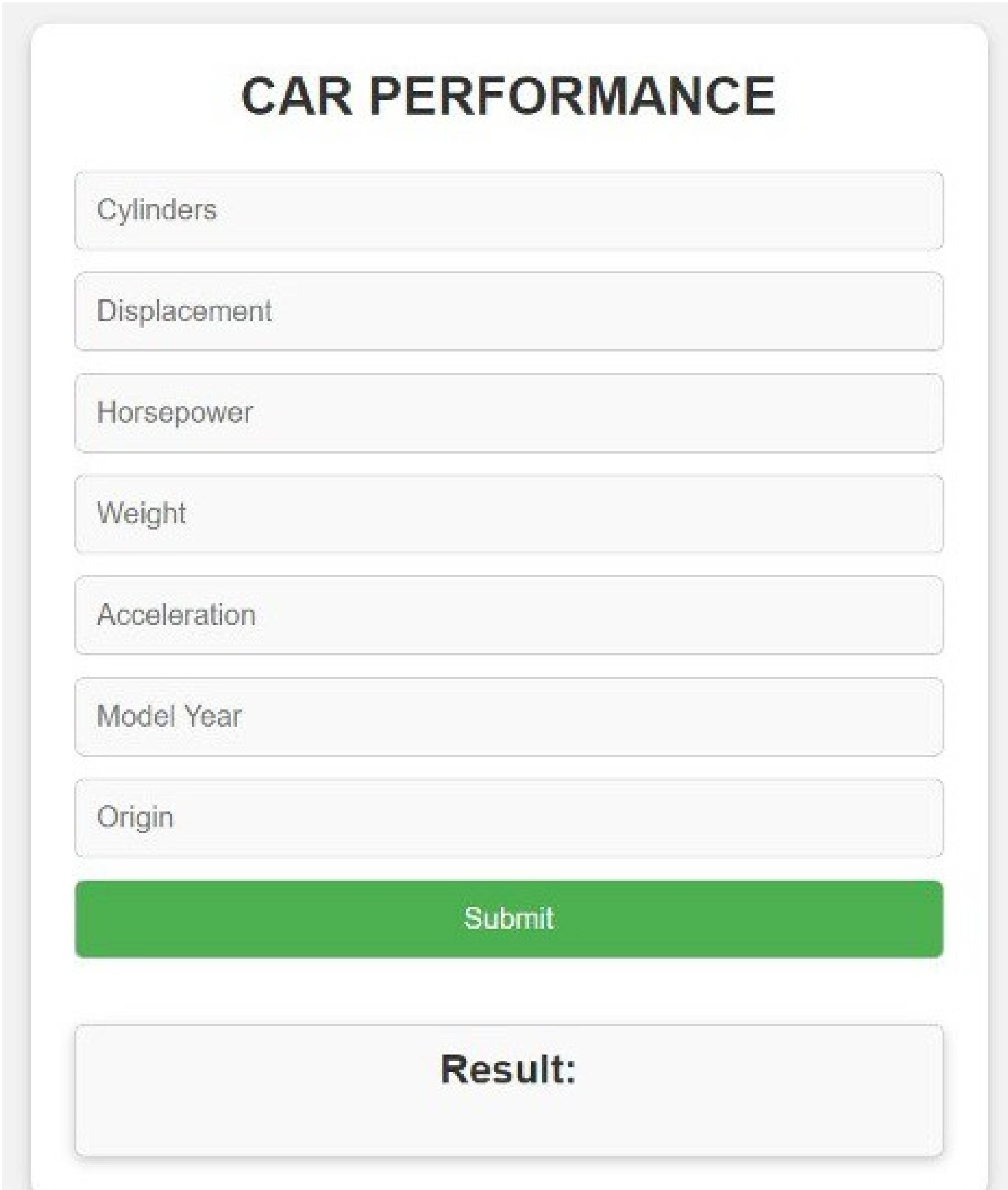
	RMSE
Multi Linear Regression	3.487866
Support Vector Regressor	5.209404
K Neighbors Regressor	4.434033
Random Forest Regressor	2.928835

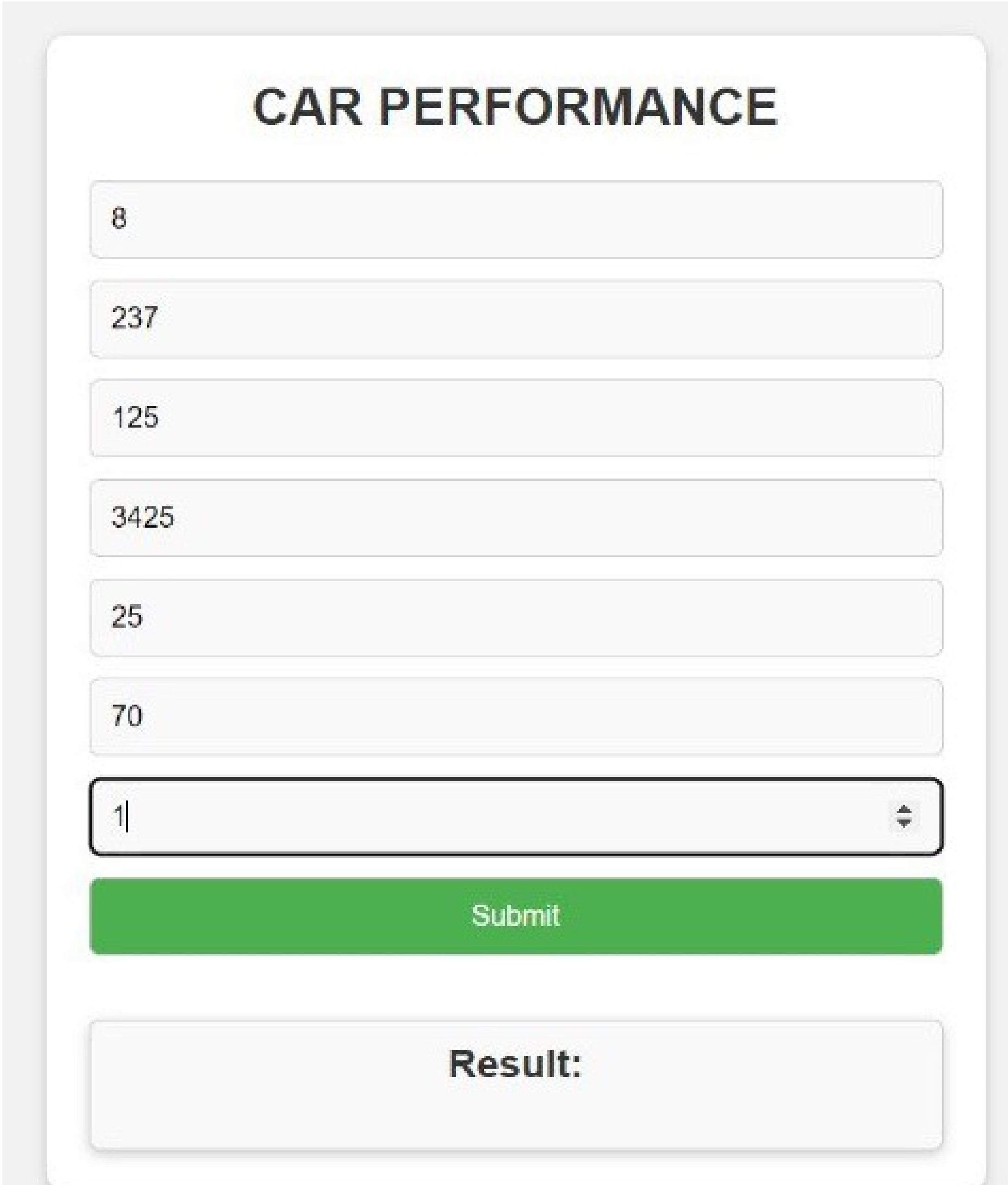
In [48]:
```
gr = df_model.plot.barh()
gr.legend(
    ncol=len(models.keys()),
    bbox_to_anchor=(0, 1),
    loc='lower left',
    prop={'size': 25}
)
plt.tight_layout()
```

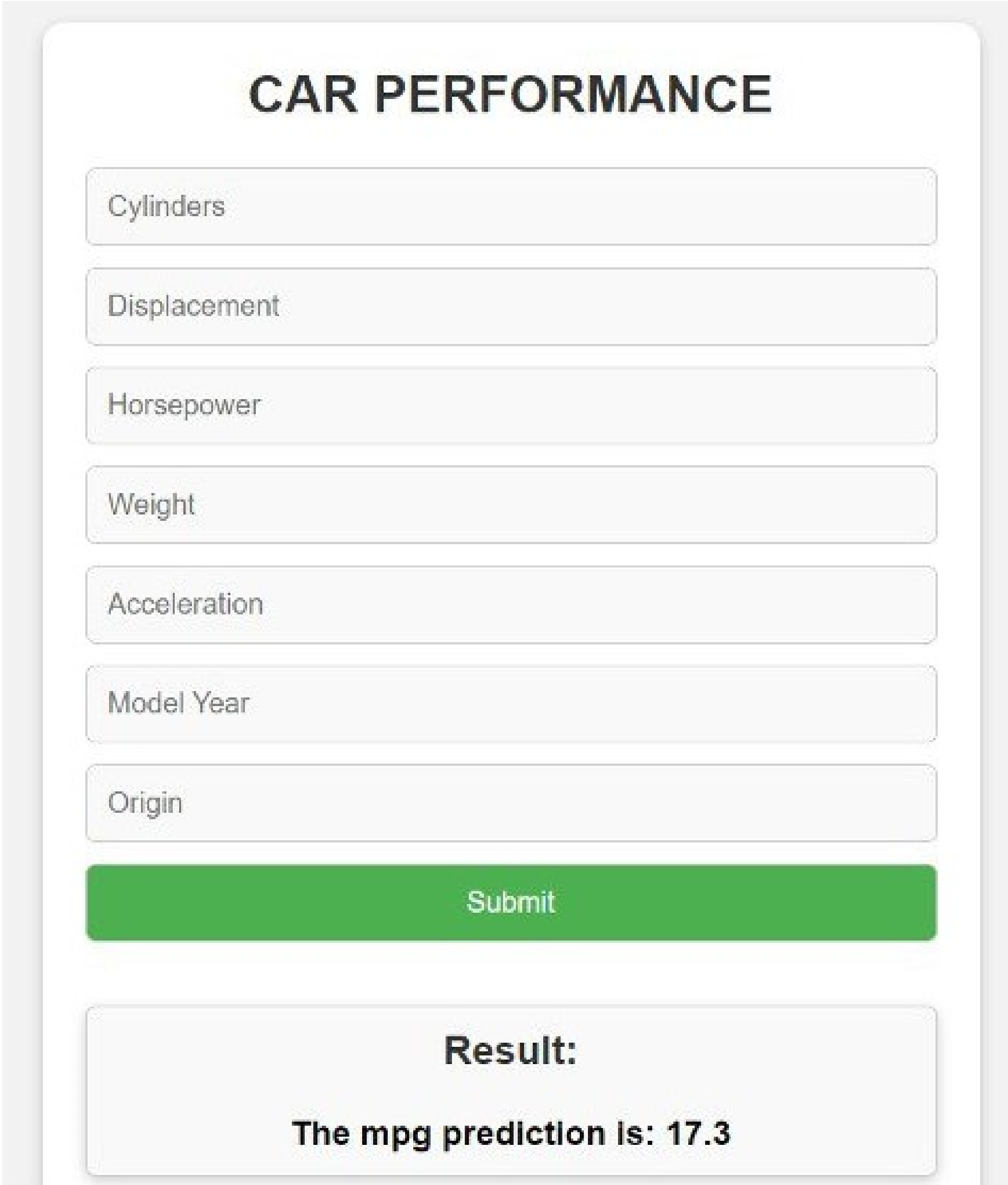


As per the above RMSE scores and graph we can see that the Random Forest Regressor has performed the best when compared to other regression models. Hence we choose Random Forest Regressor to build the model for our Car Performance Prediction Project.

The User Interface of the Web Application




CAR PERFORMANCE
8
237
125
3425
25
70
1
Submit
Result:


CAR PERFORMANCE
Cylinders
Displacement
Horsepower
Weight
Acceleration
Model Year
Origin
Submit
Result:
The mpg prediction is: 17.3

ADVANTAGES & DISADVANTAGES

Advantages

- Accurate Predictions: IBM Watson's machine learning capabilities enable accurate predictions of car performance, considering various factors and variables.
- Real-time Insights: The system can provide real-time insights into car performance, allowing for immediate adjustments or interventions to optimize performance
- Enhanced Safety: By predicting performance characteristics, drivers can make informed decisions, improving safety by adjusting driving behavior or avoiding potentially risky situations.

Disadvantages

- Data Dependency: Accurate predictions rely on the availability of high-quality and diverse data, including historical performance records, driver behavior, and environmental conditions.
- Complex Implementation: Implementing a car performance prediction system using IBM Watson requires expertise in machine learning, data analysis, and integration with the existing vehicle infrastructure.
- Cost: Developing and maintaining a car performance prediction system using IBM Watson can involve significant costs, including the required hardware, software, and ongoing updates and maintenance.

APPLICATIONS

- Car manufacturing industry
- Automotive engineering
- Racing and motorsports
- Car enthusiasts, dealerships and rental agencies
- Insurance companies
- Research and academia and autonomous vehicles

CONCLUSION

Through the acquisition and preprocessing of car performance datasets, the project allows for comprehensive analysis and feature engineering to identify the most influential factors affecting car performance. By applying appropriate feature scaling techniques, the project ensures that the predictive models can effectively handle different scales and ranges of features.

The development and evaluation of machine learning models empower car manufacturers, automotive engineers, and enthusiasts to make informed decisions during the design, development, and customization processes. The optimized models provide accurate predictions, helping improve car performance, efficiency, and the overall driving experience.

FUTURE SCOPE

The future scope of the car performance prediction project is promising. Advancements in machine learning and data analytics, coupled with the growing availability of high-quality car performance data, offer opportunities for further development. The project can be expanded to incorporate more advanced modeling techniques, including deep learning and ensemble methods, to enhance prediction accuracy. Integration with real-time data sources and sensor data from connected cars can enable real-time performance monitoring and adaptive predictions. Furthermore, incorporating interpretability techniques can provide actionable insights into the factors driving performance predictions. The project's future scope lies in continually refining and expanding the models, integrating emerging technologies, and addressing evolving industry needs to further enhance car performance prediction capabilities.Additionally involving safety and security of the users and also a fine selection of the product for the users and what they wanna purchase by having their own pre purchase analysis of a car and its data driven approach of performance

BIBLIOGRAPHY

1. https://numpy.org/doc/
2. https://pandas.pydata.org/docs/
3. https://matplotlib.org/3.5.3/api/_as_gen/matplotlib.pyplot.html
4. https://seaborn.pydata.org/
5. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html
6. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
7. https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html
8. https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html
9. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
10. https://scikit-learn.org/stable/modules/model_evaluation.html
11. https://www.researchgate.net/publication/277830894_The_Concept_to_Measure_the_Overall_Car_Performance
12. https://www.sciencedirect.com/science/article/abs/pii/S0263224117304554
13. https://www.sciencedirect.com/science/article/pii/S2666790821002007

APPENDIX

Source Code

Car_Performance_Prediction.ipynb :
https://github.com/Rixhi7/Car-Performance-Prediction-Project/blob/main/Car_Performance_Prediction.ipynb

app.py : https://github.com/Rixhi7/Car-Performance-Prediction-Project/blob/main/app.py

finalweb.html :
https://github.com/Rixhi7/Car-Performance-Prediction-Project/blob/main/finalweb.html

Project Demo Video :
https://drive.google.com/file/d/1G-I14XyI6iE9-O4z9ytEF-L0Ggu40Cfa/view?usp=sharing