MACHINE LEARNING BASED VEHICLE PERFORMANCE ANALYZER

IBM-Project-26685-1660033317

Bonafide record of work done by

Aakriti Ghimire	(1919102002)
Abhishek Purbey	(1919102003)
Hoo Dharshana Shree Nivasha S	(1919102056)
Jaya Pradhap P	(1919102061)

Faculty Mentor: R. Priyadhrashini

Bachelor of Engineering

In

Computer Science and Engineering

November 2022

Sona College of Technology, Salem, Tamil Nadu 636005

1. INTRODUCTION

1.1 Project Overview

It's a significant and intriguing challenge to predict a car's performance level. The primary objective of the current study is to forecast automobile performance to enhance specific vehicle behaviour. This can greatly reduce the fuel consumption of the system and boost its effectiveness. Analysis of the vehicle's performance depending on the kind of engine, number of cylinders, fuel type, and horsepower, among other factors. The health of the automobile may be predicted based on these variables. Obtaining, investigating, interpreting, and documenting health data based on the three elements is a continuous process. Prediction engines and engine management systems both heavily rely on performance metrics like mileage, reliability, flexibility, and cost that may be combined. To increase the performance efficiency of the vehicle, it is crucial to analyse the elements utilizing a variety of well-known machine learning methodologies, including as linear regression, decision trees, and random forests. Automobile engineering's "hot subjects" right now revolve around the power, lifespan, and range of automotive traction batteries. We also take a performance in mileage into account here. We will create the models, utilizing various techniques and neural networks, to resolve this issue. Then, we'll compare which algorithm accurately predicts car performance (mileage).

1.2 Purpose

Predicting the performance level of cars is an important and interesting problem. The main goal is to predict the performance of the car to improve certain behaviors of the vehicle. This can significantly help to improve the system's fuel consumption and increase efficiency. The performance analysis of the car is based on the engine type, no of engine cylinders, fuel type, horsepower, etc. These are the factors on which the health of the car can be predicted. It is an ongoing process of obtaining, researching, analyzing, and recording health based on these three factors. The performance objectives like mileage, dependability, flexibility and cost can be grouped together to play a vital role in the prediction engine and engine management system. This approach is a very important step towards understanding the vehicle's performance.

2. LITERATURE SURVEY

2.1 Existing problem

a) **An approach of modelling on dynamic performance evaluation for off-road vehicle:**

Automobile dynamic is one of the most important performance indexes, the key is whether the simulation accords with the real status, whether the vehicle performance under real driving conditions can be reflected more validly, and it will be used to estimate automobile dynamic accurately, or to provide theory reference for vehicle design. On the point of view of using vehicle, considering the effects of external factors as air velocity, tyre slip, adhesion coefficient on vehicle dynamic performance, a novel method on dynamic performance evaluation is established, and the effectiveness of the dynamic performance model is verified.

b) **Steering performance simulation of three-axle vehicle with multi-axle dynamic steering:**

Because three-axle heavy-vehicle with front-wheel steering has big radius at low speed and bad stability at high speed, in order to improve heavy vehicle steering performance at different speed, the multi-axle dynamic steering technology is put forward. Selecting zero side-slip angle of mass centre and proportional control strategy to control vehicle, Using MATLAB, the steering performance of the three-axle vehicle with different steering modes are simulated. The result shows that multi-axle dynamic steering can decrease the steering radius at low speed and improve vehicle stability at high speed.

c) **Simulation study on synthetical performance of electric vehicles:**
This paper presents the software development on the performance simulation of electric vehicles. Software verification is carried out via the comparison of simulation results with on-road test. Applications of the software in prototype design are also presented in terms of theoretical inference, modelling, software development and simulation of synthetical performance for EVs such as dynamic performance, economy performance as well as analysis of parameters' influences on EV performance. The commonly used European drive cycle is adopted for simulation in the paper. Simulation with the software proves an efficient and money-saving means for prototyping of EV or HEV systems with control units.

d) **Simulation and Analysis of Performance of a Pure Electric Vehicle with a Super-capacitor:**
Energy storage and power boost are major problems in the development of electric vehicles (EV). Installing a super-capacitor as an auxiliary power source to improve the performance of electric vehicles is a feasible and realistic solution. In this paper, the structure of a multi-energy system and the principles of flow of the multi energy of electric vehicles were introduced first, explaining how different sources of energy work in different situations. A model of electric vehicle with a battery and a super capacitor, based on MATLAB/Simulink was built up. The model was validated by comparing the simulation results and the actual data from later field tests. The drive cycle used in the simulation was the CYC_CONST_45. Comparisons were made between the model of the vehicle with a super capacitor and the model of a vehicle without a super-capacitor. Field tests of an electric vehicle were conduct and analyses were made. The analysis includes vehicle dynamic performance and economic performance in urban environments where the vehicle accelerated and decelerated frequently. The results showed that installing a super-capacitor improves the working conditions of the battery. The variation of the current drawn by the vehicle was smoothed due to the working of the super-capacitor, which provided better working conditions for the battery and increased the operating life of the battery.

e) **Steering feel study on the performance of EPS:**
The steering feel study is very important in the development of electric power steering system (EPS). This paper describes a method about how to evaluate and get the suitable steering feel when driving a vehicle equipped with EPS. The EPS steering feel subjective tests were performed to obtain objective quality parameters that correlate with subjective evaluation. After this, the paper briefly describes the statistical technique used to identify which parameters best correlate with vehicle steering qualities. As there was no correlation between a single partial rating and a single objective indicator, the principal component analysis (PCA) method was chosen and obtained objective indices. The objective evaluation parameters have been validated by drivers' subjective evaluation. In the third part, the analytical method was applied to vehicle dynamic analysis to analyse vehicle steering feel characteristics, we established a closed-loop steering feel simulation model to analyse steering torque characteristics, vehicle dynamic response and assess steering feel performance for different settings of a EPS system. The design of EPS was optimized and achieved more suitable driving feel by using the dynamic analysis model without plenty of real vehicle tests. This method makes it possible to easily and accurately benchmark steering dynamic characteristics, set design targets, and is helpful to achieve good steering feel.

f) **Study on the performance and control of SR machine for vehicle regenerative braking:**
A regenerative braking system with simple structure, high efficiency, good performance and easy

control is crucial for electric vehicle (EV), hybrid electric vehicle (HEV) and fuel cell vehicle (FCV). SR machine is one of the promising candidates. In this paper, the current and torque performance of a SR machine for application to vehicle regenerative balding has been studied. The relationship between the torque, speed, turn-on and turn-off angles has been established. The data obtained through simulation is very useful for vehicle control design.

g) **A method to analyse driver influence on the energy consumption and power needs of electric vehicles:**
The energy consumption and power needs of electric vehicles are evaluated on roller test benches according to test procedures defined by legal standards and by vehicle manufacturers. These test procedures are mainly defined by driving cycles and include tolerances to compensate for the human error during these tests. These tolerances may seem to make the tests easier but they can have a big effect on the appropriate dimensioning of the components, and also on the performance of the vehicle. Within this paper, a method is presented, which enables the quantification of these effects depending on the type of the test procedure, and the way the driving cycle is driven. The developed method has been tested in a simulation environment and several standard test procedures were analysed.

h) **Autonomous underwater vehicle minimum-time navigation in a current field:**
The problem of navigation in a spatially variable current is reviewed, and for a certain class of mathematically-describable functions, solved for minimum time in closed form.

i) **Transmission system performance analysis of traditional power vehicle:**
Based on simulation software GT-drive, the author analysed the transmission system performance of a passenger car with diesel engine and provided the appropriate research methods. Firstly, the numerical simulation model of a vehicle was built based on vehicle weight, frontal area, rolling, air-drag coefficient, etc. The different matching schemes were simulated and compared. The results show that, for a given engine, using different transmission systems, the matching efficiency is significantly different. In view of power and economy of the vehicle, it is important that selected suitable power transmission device. This method has provided a theoretical basis for studying traditional power vehicle, also giving some information to study the new type vehicle power train system.

j) **Real-time performance of control allocation for actuator coordination in heavy vehicles:**
This paper shows how real-time optimisation for actuator coordination, known as control allocation, can be a viable choice for heavy vehicle motion control systems. For this purpose, a basic stability control system implementing the method is presented. The real-time performance of two different control allocation solvers is evaluated and the use of dynamic weighting is analysed. Results show that sufficient vehicle stability can be achieved when using control allocation for actuator coordination in heavy vehicle stability control. Furthermore, real-time simulations indicate that the optimisation can be performed with the computational capacity of today's standard electronic control units.

2.2 References

a) Junshu Han, Zhenhai Gao, Shulin Tan, Xiangdong Cui, Institute of Medical Equipment, Academy of Military Medical Sciences, Tianjin, China.
Published in: 2010 8th World Congress on Intelligent Control and Automation

b) Shufeng Wang, Junyou Zhang, Huashi Li, College of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, China.
Published in: 2008 IEEE Vehicle Power and Propulsion Conference

c) Sun Fengchun, Sun Liqiug, Zhu Jiaguang, Electric Vehicle Research and Development Center, Beijing Institute of Technology, Beijing, China.

Published in: Proceedings of the IEEE International Vehicle Electronics Conference (IVEC'99) (Cat. No.99EX257)

d) N Jinrui, W Zhif, School of mechanical and vehicular engineering, Beijing Institute of Technology, Beijing, China.
Published in: 2006 IEEE Vehicle Power and Propulsion Conference

e) Xin. Zhang, Zhang Xin, School of Mechanical, Electric and Control Engineering, Beijing Jiaotong University, Beijing, China. Shi Guobiao, Electric Vehicle Center of Analysis and Technology, Beijing Institute of Technology, Beijing, China.
Published in: 2008 IEEE Vehicle Power and Propulsion Conference

f) Xiaoling Yuan, College of Electrical Engineering, Hohai University, HHU, Nanjing, Jiangsu, China. Yimin Gao, M. Ehsani, Department of Electrical Engineering, Texas A and M University, College Station, TX, USA.
Published in: 2008 IEEE Vehicle Power and Propulsion Conference

g) Rayad Kubaisi, Frank Gauterin, Martin Giessler, Chair of Vehicle Technology, Institute of Vehicle System Technology, Karlsruhe, Germany.
Published in: 2009 IEEE Intelligent Symposium

h) Max Blanco, Philip A. Wilson, Fluid-Structure Interactions Group, School of Engineering Sciences, University of Southampton, Southampton, UK.
Published in: OCEANS 2010 MTS/IEEE SEATTLE

i) Feng Kang, Liu Jingping, Fu Jianqin, Yang Hanqian, Research Center of Advanced Powertrain Technology, State Key Laboratory of Advanced D&M for Vehicle Body, Hunan University, Changsha, China.
Published in: 2011 International Conference on Electric Information and Control Engineering

j) Kristoffer Tagesson, Leo Laine, Department Chassis Strategies & Vehicle Analysis, VOLVO 3PDepartment 26661, AB4S GOTEBORG, Sweden. Peter Sundstrom, MaDELON AB, Lund, Sweden, Nicolas Dela.
Published in: 2009 IEEE Intelligent Vehicles Symposium

2.3 Problem Statement Definition

Abstract

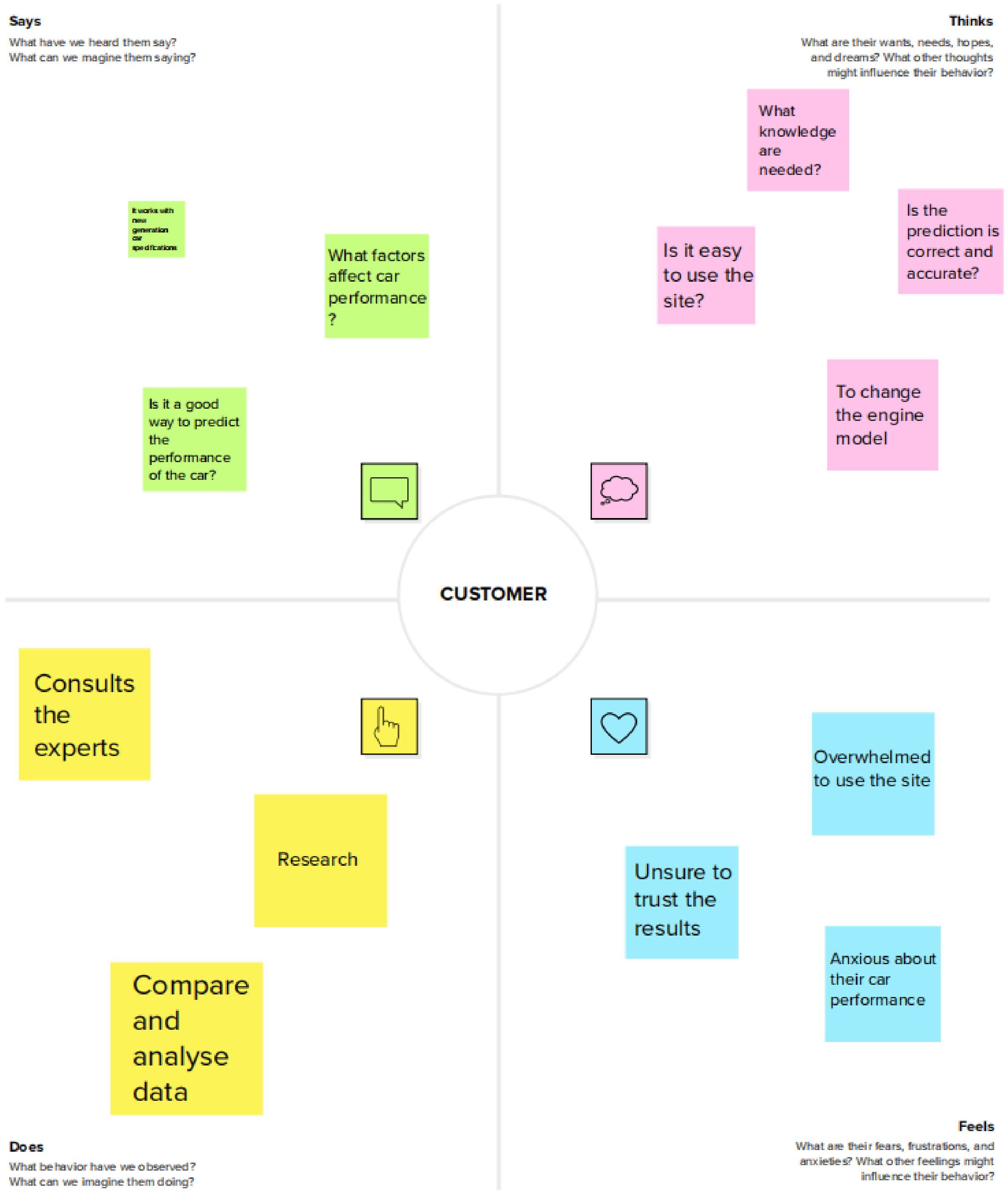
Automobile manufacturer or a user, who is looking for new technologies to dominate the extremely competitive automobile market. The problem affects automobile companies struggling with improving their vehicle performance. Automobile manufacturing & testing process; Automobile repair & service process; People who wish to know the performance of their vehicles. Requirement to analyse the vehicle fuel economy, integrated safety, drivability, durability, features, aerodynamic performance, etc. Increasing fuel prices forces vehicles to have higher fuel efficiency for becoming a market hit. Automobiles require constant introduction of new features to survive in the rapidly expanding auto mobile industry. People prefer maintenance of old vehicles than buying new vehicle, which require automobiles to provide better and easier maintenance of their vehicles. People feel better when they are constantly updated on their vehicle performance, vehicle safety standards, etc. Automobile manufacturing companies, Automobile service stations, People, etc.

Vehicle performance analyser does not only benefit the automobile companies or the people using it. Improving vehicle performance by analysing the vehicle, leads to betterment of the vehicles in various aspects. For instance, improving fuel efficiency conserves fuel, reducing the emissions reduces the stress on the environment, etc.

3. IDEATION & PROPOSED SOLUTION

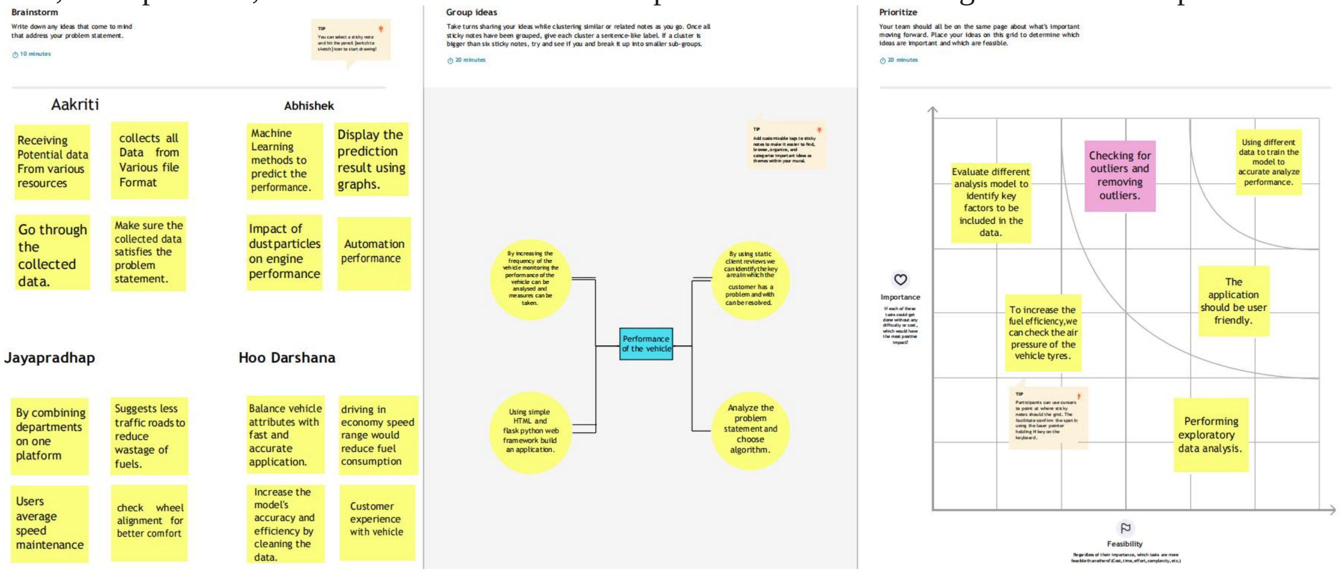
3.1 Empathy Map

The primary purpose of the empathy map is to bridge the understanding of the user and developer.


Says
What have we heard them say?
What can we magine them saying?
It works with new generation car specifications
What factors affect car performance ?
Is it a good way to predict the performance of the car?
Thinks
What are their wants, needs, hopes, and dreams? What other thoughts might influence their behavior?
What knowledge are needed?
Is it easy to use the site?
Is the prediction is correct and accurate?
To change the engine model
CUSTOMER
Consults the experts
Research
Compare and analyse data
Does
What behavior have we observed?
What can we imagine them doing?
Overwhelmed to use the site
Unsure to trust the results
Anxious about their car performance
Feels
What are their fears, frustrations, and anxieties? What other feelings might influence their behavior?

3.2 Ideation & Brainstorming

The aim is to generate a large quantity of ideas that the team can then filter and cut down into the best, most practical, or most innovative ones to inspire new and better design solutions and products.

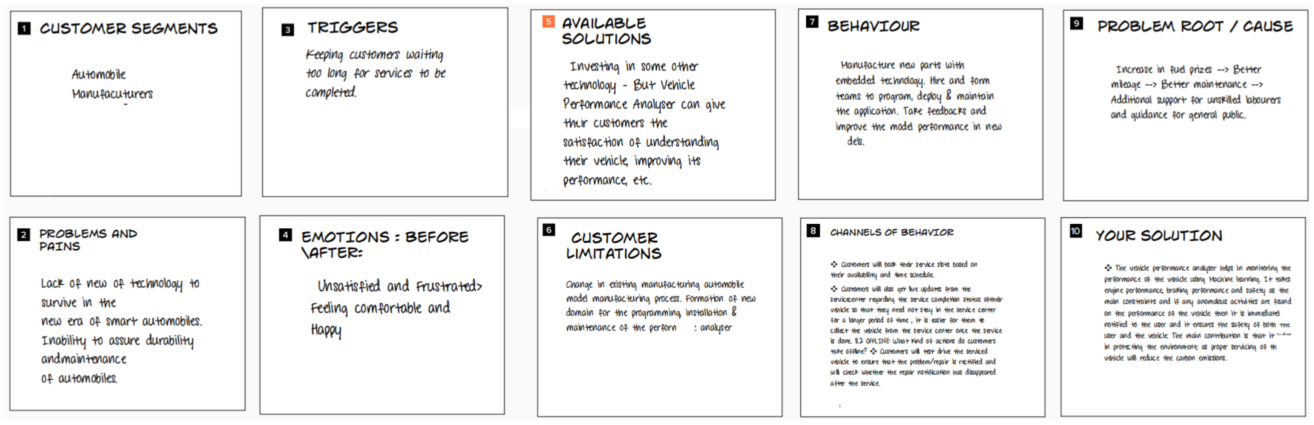

Brainstorm
Write down any ideas that come to mind that address your problem statement.
10 minutes
TIP
You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!
Aakriti
Receiving Potential data From various resources
collects all Data from Various file Format
Go through the collected data.
Make sure the collected data satisfies the problem statement.
Abhishek
Machine Learning methods to predict the performance.
Display the prediction result using graphs.
Impact of dust particles on engine performance
Automation performance
Jayapradhap
By combining departments on one platform
Suggests less traffic roads to reduce wastage of fuels.
Users average speed maintenance
check wheel alignment for better comfort
Hoo Darshana
Balance vehicle attributes with fast and accurate application.
driving in economy speed range would reduce fuel consumption
Increase the model's accuracy and efficiency by cleaning the data.
Customer experience with vehicle
Group ideas
Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.
20 minutes
TIP
Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.
By increasing the frequency of the vehicle monitoring the performance of the vehicle can be analysed and measures can be taken.
By using static client reviews we can identify the key area in which the customer has a problem and with can be resolved.
Performance of the vehicle
Using simple HTML and flask python web framework build an application.
Analyze the problem statement and choose algorithm.
Prioritize
Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.
20 minutes
Evaluate different analysis model to identify key factors to be included in the data.
Checking for outliers and removing outliers.
Using different data to train the model to accurate analyze performance.
Importance
If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?
To increase the fuel efficiency,we can check the air pressure of the vehicle tyres.
The application should be user friendly.
TIP
Participants can use cursors to point at where sticky notes should the grid. The facilitate confirm the spot by using the laser pointer holding H key on the keyboard.
Performing exploratory data analysis.
Feasibility
Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)

3.3 Proposed Solution

S.no.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Predicting the performance level of cars is an important and interesting problem. The main goal is to predict the performance of the car to improve certain behaviours of the vehicle. This can significantly help to improve the system's fuel consumption and increase efficiency. The performance analysis of the car is based on the engine type, no of engine cylinders, fuel type, horsepower, etc. These are the factors on which the health of the car can be predicted. It is an on-going process of obtaining, researching, analysing, and recording health based on the above three factors. The performance objectives like mileage, dependability, flexibility and cost can be grouped together to play a vital role in the prediction engine and engine management system. This approach is a very important step towards understanding the vehicle's performance.
2.	Idea/Solution description	To train the system with the dataset using a regression model and it will be integrated to the web-based application where the user is notified with the status.
3.	Novelty/ Uniqueness	Giving the public and the manufacturer the feature to analyse their vehicle’s performance.
4.	Social Impact / Customer Satisfaction	The petrol/diesel cost can become lower due to a better mileage performance and the existing vehicle parts can be reused which increases the re-usability thus decreases the cost on new products and the physically able people have better seat comfort because of accessories work. Better mileage and better engine maintenance provides complete combustion thus emitting less harmful gases.

5.	Business Model (Revenue Model)	The web-based application has a friendly UI for the customer to enter their vehicles detail and the system predicts the value within few seconds.
6.	Scalability of the Solution	The project will be scalable when the parts used to measure data in vehicles is feasible and the ML model is fast in processing data.

3.4 Problem Solution fit



1 CUSTOMER SEGMENTS

Automobile Manufacuturers

3 TRIGGERS

Keeping customers waiting too long for services to be completed.

5 AVAILABLE SOLUTIONS

Investing in some other technology - But Vehicle Performance Analyser can give their customers the satisfaction of understanding their vehicle, improving its performance, etc.

7 BEHAVIOUR

Manufacture new parts with embedded technology. Hire and form teams to program, deploy & maintain the application. Take feedbacks and improve the model performance in new dels.

9 PROBLEM ROOT / CAUSE

Increase in fuel prizes --> Better mileage --> Better maintenance --> Additional support for unskilled labourers and guidance for general public.

2 PROBLEMS AND PAINS

Lack of new of technology to survive in the new era of smart automobiles. Inability to assure durability andmaintenance of automobiles.

4 EMOTIONS : BEFORE \AFTER:

Unsatisfied and Frustrated> Feeling comfortable and Happy

6 CUSTOMER LIMITATIONS

Change in existing manufacturing automobile model manufacturing process. Formation of new domain for the programming, installation & maintenance of the perforn : analyser

8 CHANNELS OF BEHAVIOR

❖ Customers will book their service slots based on their availability and time schedule

❖ Customers will also get live updates from the servicecenter regarding the service completion status oftheir vehicle so that they need not stay in the service center for a longer period of time , it is easier for them to collect the vehicle from the service center once the service is done. 8.2 OFFLINE What kind of actions do customers take offline? ❖ Customers will test drive the serviced vehicle to ensure that the problem/repair is rectified and will check whether the repair notification has disappeared after the service.

10 YOUR SOLUTION

❖ The vehicle performance analyser helps in monitoring the performance of the vehicle using Machine learning. It takes engine performance, braking performance and safety as the main constraints and if any anomalous activities are found on the performance of the vehicle then it is immediatel notified to the user and it ensures the safety of both the user and the vehicle. The main contribution is that it helps in protecting the environment; as proper servicing of th vehicle will reduce the carbon emissions.

4. REQUIREMENT ANALYSIS

4.1 Functional requirements

Fr no.	Functional requirement (epic)	Sub requirement (story / sub-task)
FR-1	User Registration	Registration through Website
FR-2	Data processing	Process raw data and perform manual analysis
FR-3	Model building	Get the predicted performance of the vehicle using the given data
FR-4	Results of the analysis	Displaying the fuel consumption by vehicle

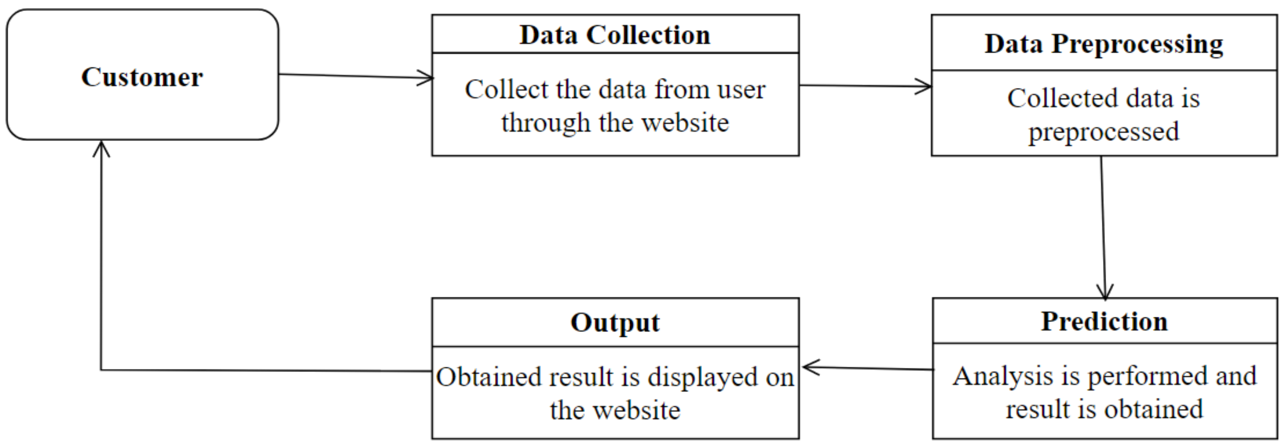
4.2 Non-Functional requirements

NFR No.	Non-Functional Requirement	Description
NFR-1	Usability	Useful for automobile engineers and customers who want to purchase new car
NFR-2	Security	Providing the website with basic security
NFR-3	Reliability	Providing accurate performance values for the vehicle
NFR-4	Performance	Improving performance by utilizing an efficient ML algorithm
NFR-5	Availability	Available to anyone and anytime through Internet
NFR-6	Scalability	Will be able to handle multiple users at a time

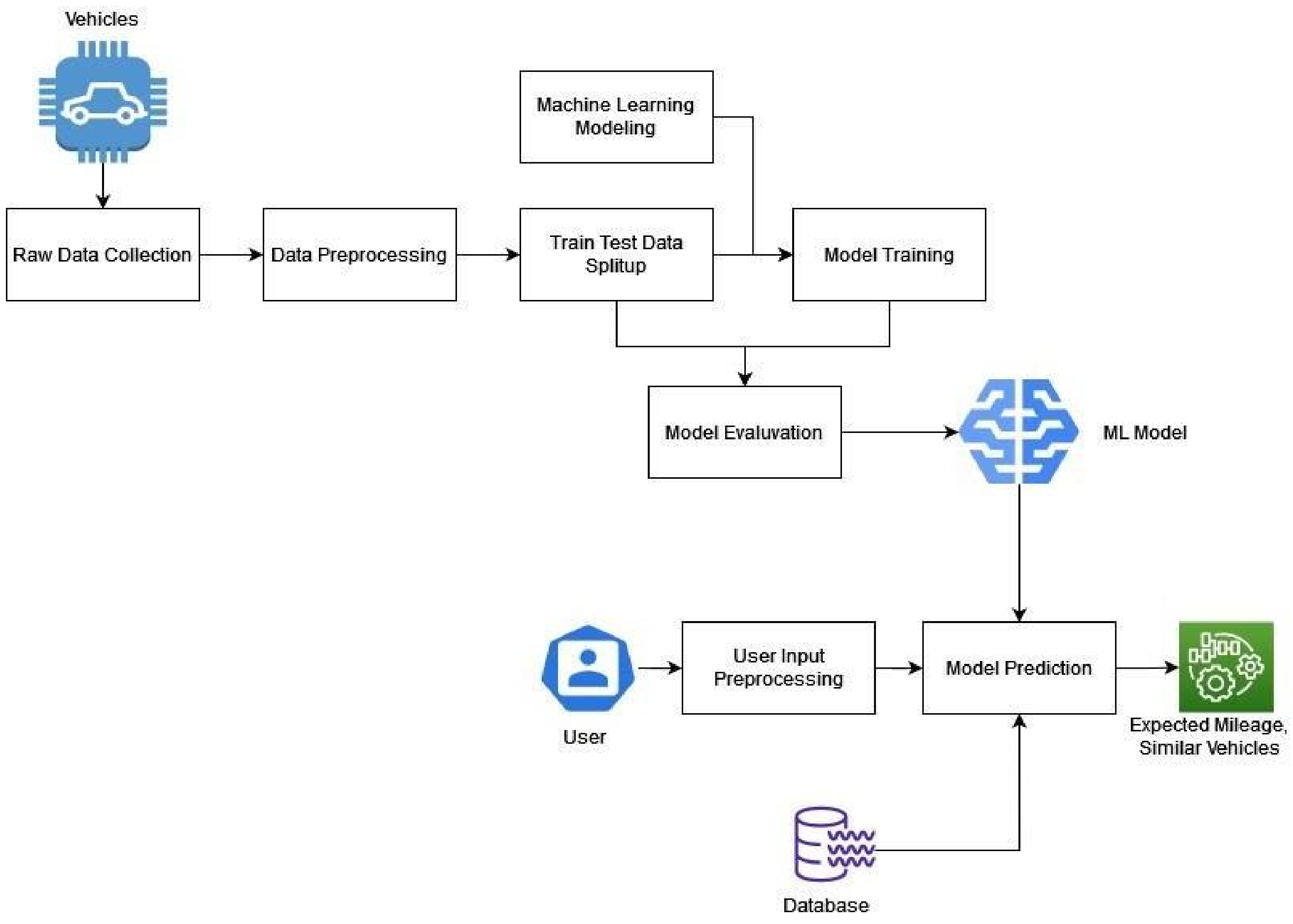
5. PROJECT DESIGN

5.1 Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of how information flows within a system. A neat and clear DFD can thus depict the right amount of the system requirements graphically. It not only shows how data enters and leaves the system, but also what changes the information and where the data is stored.


Customer
Data Collection
Collect the data from user through the website
Data Preprocessing
Collected data is preprocessed
Prediction
Analysis is performed and result is obtained
Output
Obtained result is displayed on the website

5.2 Solution & Technical Architecture



5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story/Task	Acceptance criteria	Priority	Release
Customer (Desktop user)	Input	USN-1	User can enter the car details in the website.	Enters car details.	High	Sprint-1
Customer (Desktop User)	Output	USN-2	User can obtain performance metrics and the analysis from the website.	Obtain the results.	High	Sprint-4

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data processing	USN-1	As a user, I can process raw data and perform manual analysis.	20	High	Aakriti Abhishek Hoo Dharshana Jayapradhap
Sprint-2	Model building	USN-2	As a user, I can get the predicted performance of the vehicle using the given data.	20	Low	Aakriti Abhishek Hoo Dharshana Jayapradhap
Sprint-3	Web Page design	USN-3	As a user, I am able to view the website and I can get the predicted performance of the vehicle using the given data.	20	High	Aakriti Abhishek Hoo Dharshana Jayapradhap
Sprint-4	Result	USN-4	As a user, I expect the prediction is highly accurate.	20	High	Aakriti Abhishek Hoo Dharshana Jayapradhap

6.2 Sprint Delivery Schedule

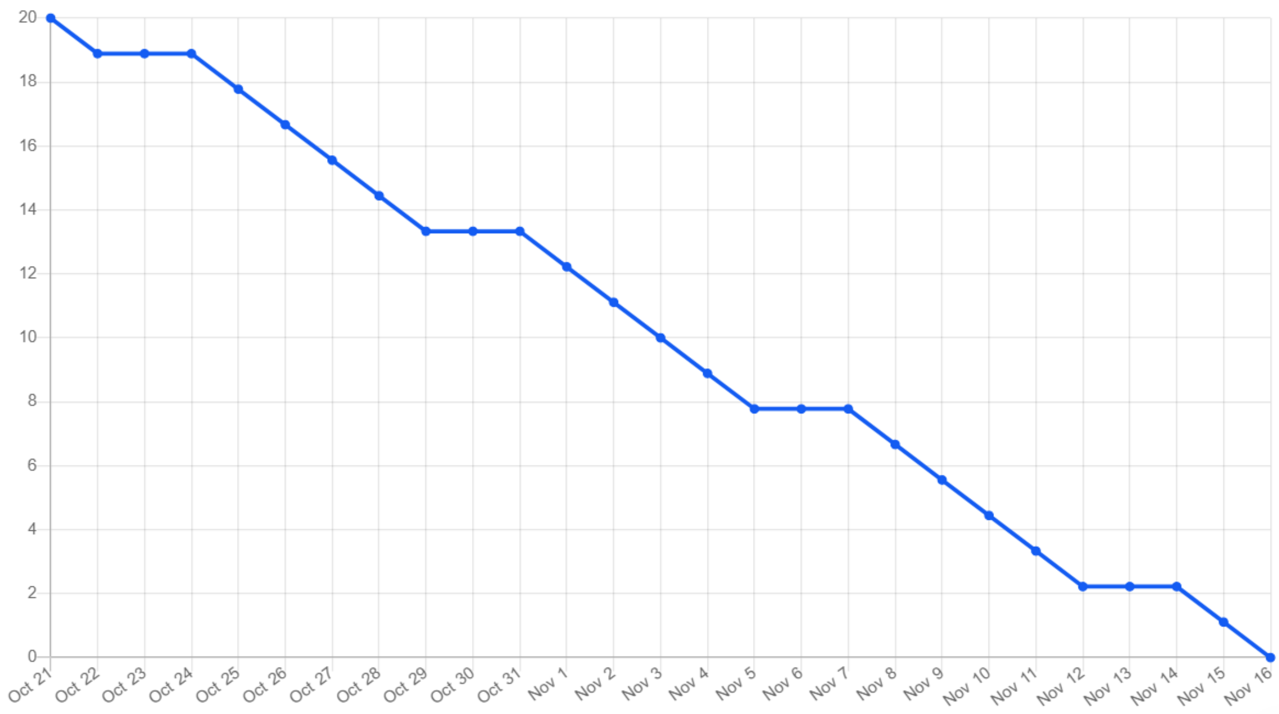
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date	Story Points Completed	Sprint Release Date
Sprint-1	20	6 days	21 Oct 2022	26 Oct 2022	20	26 Oct 2022
Sprint-2	20	6 days	28 Oct 2022	03 Nov 2022	20	03 Nov 2022
Sprint-3	20	6 days	04 Nov 2022	09 Nov 2022	20	09 Nov 2022
Sprint-4	20	6 days	11 Nov 2022	16 Nov 2022	20	16 Nov 2022

6.3 Reports from JIRA

Velocity:
Imagine we have a 6-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day).

$$\text{Average Velocity} = \frac{sprint\ duration}{velocity} = \frac{20}{6} = 3.33$$

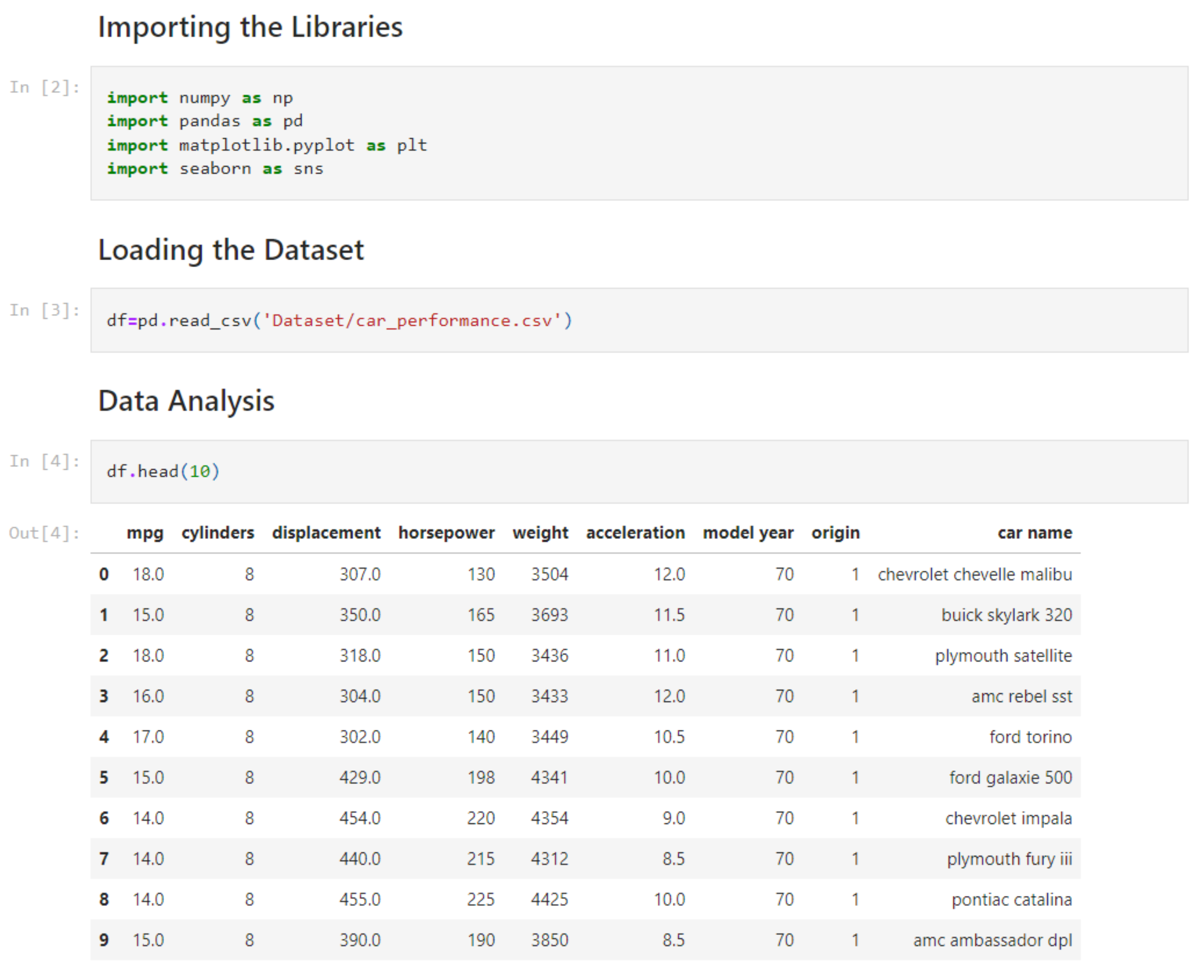
Burndown Chart:
A burndown chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.



7. CODING & SOLUTIONING

7.1 Feature 1

Data Pre-processing



Importing the Libraries

In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading the Dataset

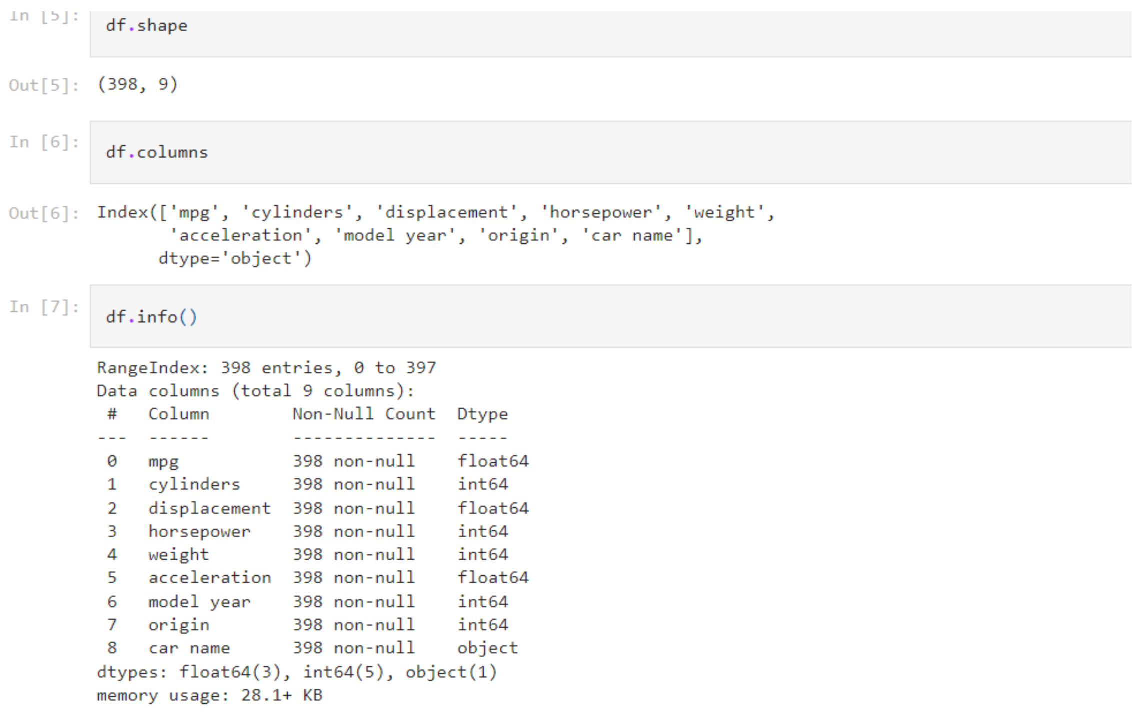
In [3]:
```
df=pd.read_csv('Dataset/car_performance.csv')
```

Data Analysis

In [4]:
```
df.head(10)
```

Out[4]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	car name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino
5	15.0	8	429.0	198	4341	10.0	70	1	ford galaxie 500
6	14.0	8	454.0	220	4354	9.0	70	1	chevrolet impala
7	14.0	8	440.0	215	4312	8.5	70	1	plymouth fury iii
8	14.0	8	455.0	225	4425	10.0	70	1	pontiac catalina
9	15.0	8	390.0	190	3850	8.5	70	1	amc ambassador dpl



In [5]:
```
df.shape
```

Out[5]:
```
(398, 9)
```

In [6]:
```
df.columns
```

Out[6]:
```
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
       'acceleration', 'model year', 'origin', 'car name'],
      dtype='object')
```

In [7]:
```
df.info()
```

```
RangeIndex: 398 entries, 0 to 397
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   mpg           398 non-null    float64
 1   cylinders     398 non-null    int64
 2   displacement  398 non-null    float64
 3   horsepower    398 non-null    int64
 4   weight        398 non-null    int64
 5   acceleration  398 non-null    float64
 6   model year    398 non-null    int64
 7   origin        398 non-null    int64
 8   car name      398 non-null    object
dtypes: float64(3), int64(5), object(1)
memory usage: 28.1+ KB
```

In [11]:
```
df.nunique()
```

Out[11]:
```
mpg             129
cylinders         5
displacement     82
horsepower       93
weight          351
acceleration     95
model year       13
origin            3
car name        305
dtype: int64
```

In [13]:
```
df.origin.unique()
```

Out[13]:
```
array([1, 3, 2])
```

Handiling the Missing Values

In [15]:
```
df.isna().sum()
```

Out[15]:
```
mpg             0
cylinders       0
displacement    0
horsepower      0
weight          0
acceleration    0
model year      0
origin          0
car name        0
dtype: int64
```

In [16]:
```
# There is no Null Value in the data set
```

Lable encoding

In [17]:
```
# There is no Categorial value other than the car name (car name is not used for the performance predecting so we can drop the car name column), so we
```

Droping the car name column

In [18]:
```
df=df.iloc[:,:-1]
```

In [19]:
```
df.head()
```

Out[19]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin
0	18.0	8	307.0	130	3504	12.0	70	1
1	15.0	8	350.0	165	3693	11.5	70	1
2	18.0	8	318.0	150	3436	11.0	70	1
3	16.0	8	304.0	150	3433	12.0	70	1
4	17.0	8	302.0	140	3449	10.5	70	1

Splitting the dataset into dependent and independent Variable

In [20]:
```
x=df.iloc[:,1:]
```

In [21]:
```
y=df.iloc[:,0]
```

In [23]:
```
x.head()
```

Out[23]:

	cylinders	displacement	horsepower	weight	acceleration	model year	origin
0	8	307.0	130	3504	12.0	70	1
1	8	350.0	165	3693	11.5	70	1
2	8	318.0	150	3436	11.0	70	1
3	8	304.0	150	3433	12.0	70	1
4	8	302.0	140	3449	10.5	70	1

In [24]:
```
y.head()
```

```
Out[24]: 0    18.0
         1    15.0
         2    18.0
         3    16.0
         4    17.0
         Name: mpg, dtype: float64
```

Splitting the dataset into train and test

```
In [25]: from sklearn.model_selection import train_test_split
         x_train,x_test, y_train, y_test = train_test_split(x,y,test_size=0.2)
```

```
In [26]: x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

```
Out[26]: ((318, 7), (80, 7), (318,), (80,))
```

Normalizing the values

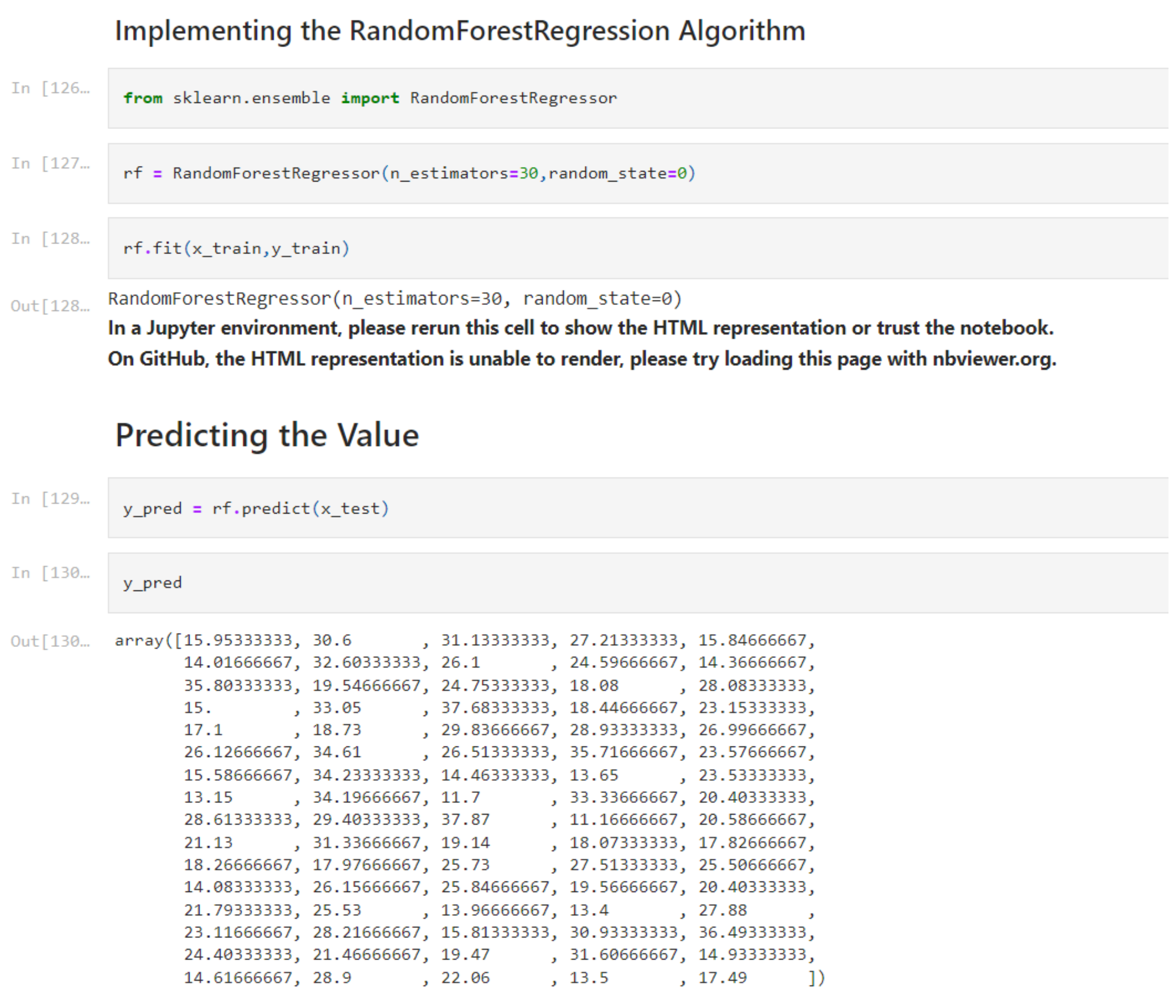
```
In [28]: from sklearn.preprocessing import StandardScaler
         sd = StandardScaler()
         x_train=sd.fit_transform(x_train)
         x_test=sd.fit_transform(x_test)
```

```
In [30]: x_train
```

```
Out[30]: array([[ 0.32894571, -0.34956192,  0.47636441, ..., -0.74142165,
                 -0.81838932,  1.77992292],
                [ 0.32894571,  0.07155568, -0.49772381, ...,  0.95037804,
                  0.832231  , -0.71904171],
                [-0.85302871, -0.50269559, -0.36609027, ..., -0.02150689,
                 -1.36859609, -0.71904171],
                ...,
                [ 0.32894571,  0.55009841,  0.02881036, ..., -0.38146427,
                  0.00692084, -0.71904171],
                [-0.85302871, -0.98123832, -0.7609909 , ..., -0.38146427,
                 -0.54328593, -0.71904171],
                [-0.85302871, -0.90467148, -0.94527786, ...,  1.05836525,
                  0.28202423,  1.77992292]])
```

```
In [ ]:
```

Model Building



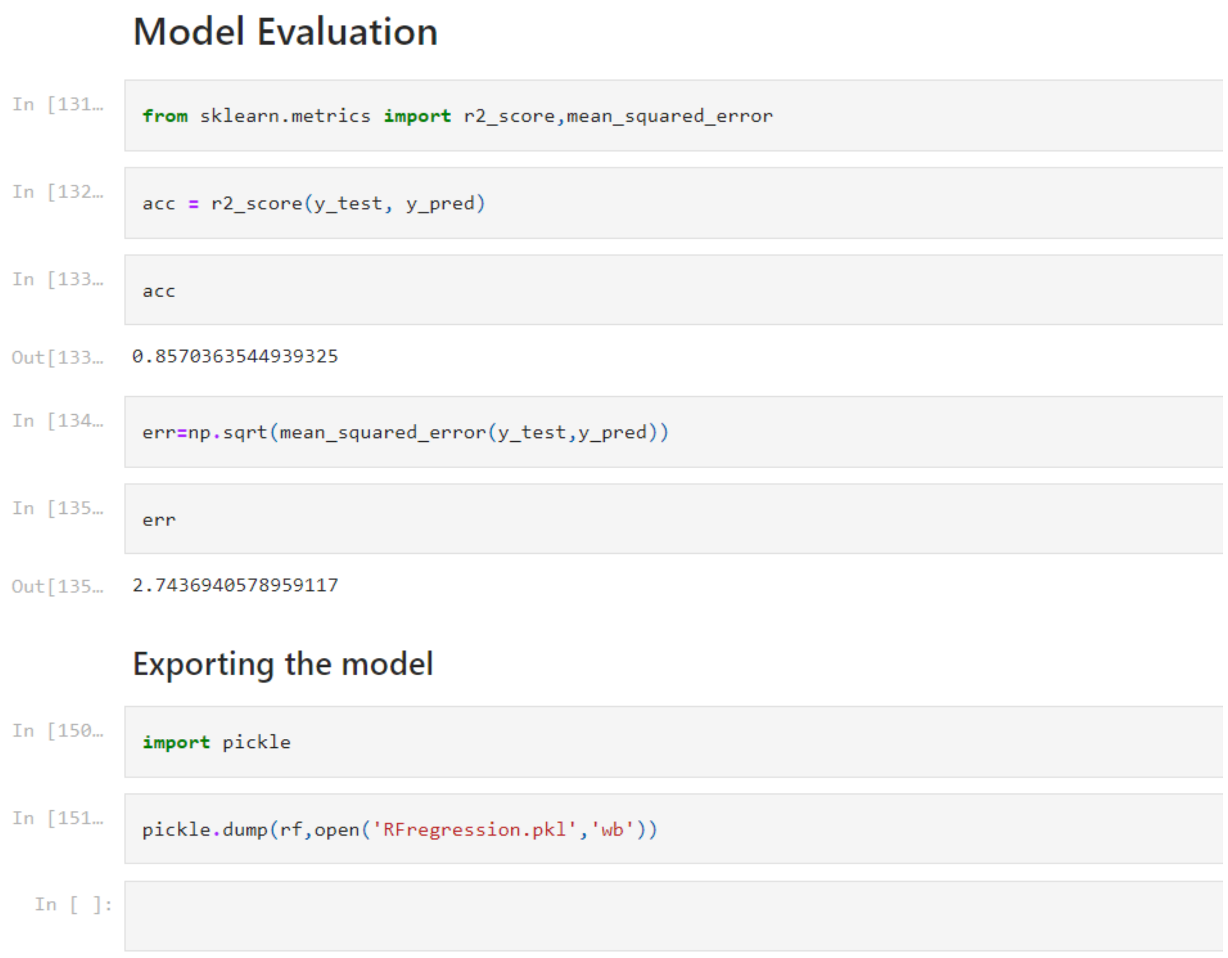
Implementing the RandomForestRegression Algorithm

```
In [126…  from sklearn.ensemble import RandomForestRegressor

In [127…  rf = RandomForestRegressor(n_estimators=30,random_state=0)

In [128…  rf.fit(x_train,y_train)

Out[128…  RandomForestRegressor(n_estimators=30, random_state=0)
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

Predicting the Value

```
In [129…  y_pred = rf.predict(x_test)

In [130…  y_pred

Out[130…  array([15.95333333, 30.6       , 31.13333333, 27.21333333, 15.84666667,
                14.01666667, 32.60333333, 26.1       , 24.59666667, 14.36666667,
                35.80333333, 19.54666667, 24.75333333, 18.08      , 28.08333333,
                15.        , 33.05      , 37.68333333, 18.44666667, 23.15333333,
                17.1       , 18.73      , 29.83666667, 28.93333333, 26.99666667,
                26.12666667, 34.61      , 26.51333333, 35.71666667, 23.57666667,
                15.58666667, 34.23333333, 14.46333333, 13.65      , 23.53333333,
                13.15      , 34.19666667, 11.7       , 33.33666667, 20.40333333,
                28.61333333, 29.40333333, 37.87      , 11.16666667, 20.58666667,
                21.13      , 31.33666667, 19.14      , 18.07333333, 17.82666667,
                18.26666667, 17.97666667, 25.73      , 27.51333333, 25.50666667,
                14.08333333, 26.15666667, 25.84666667, 19.56666667, 20.40333333,
                21.79333333, 25.53      , 13.96666667, 13.4       , 27.88      ,
                23.11666667, 28.21666667, 15.81333333, 30.93333333, 36.49333333,
                24.40333333, 21.46666667, 19.47      , 31.60666667, 14.93333333,
                14.61666667, 28.9       , 22.06      , 13.5       , 17.49      ])
```



Model Evaluation

```
In [131…  from sklearn.metrics import r2_score,mean_squared_error

In [132…  acc = r2_score(y_test, y_pred)

In [133…  acc

Out[133…  0.8570363544939325

In [134…  err=np.sqrt(mean_squared_error(y_test,y_pred))

In [135…  err

Out[135…  2.7436940578959117
```

Exporting the model

```
In [150…  import pickle

In [151…  pickle.dump(rf,open('RFregression.pkl','wb'))

In [ ]:
```

7.2 Feature 2

HTML code:

```
<!DOCTYPE html>
<html>
<head>
<title>Vehicle Performance Analyzer</title>
<meta charset="utf-8">
<link rel="stylesheet" href="style.css">
</head>
<body>
<header>
<h1>Vehicle Performance Analyzer</h1>
</header>
<div class="form-container">
<form action="/model" method="POST">
<div class="field">
<label for="no_of_cylinders">
<p class="cylinders field-name">Number of Cylinders</p>
</label>
<input type="number" id="no_of_cylinders input" name="no_of_cylinders">
</div>

<div class="field">
<label for="displacement">
<p class="displacement field-name">Displacement</p>
</label>
<input type="number" id="displacement input" name="displacement">
</div>

<div class="field">
<label for="horsepower">
<p class="horsepower field-name">Horse Power</p>
</label>
<input type="number" id="horsepower input" name="horsepower">
</div>
```

```
<div class="field">
<label for="weight">
<p class="weight field-name">Weight</p>
</label>
<input type="number" id="weight input" name="weight">
</div>

<div class="field">
<label for="acceleration">
<p class="acceleration field-name">Acceleration</p>
</label>
<input type="number" id="acceleration input" name="acceleration">
</div>

<div class="field">
<label for="model_year">
<p class="model_year field-name">Model Year</p>
</label>
<input type="number" id="model_year input" name="model_year">
</div>

<div class="field">
<label for="origin">
<p class="origin field-name">Origin</p>
</label>
<input type="number" id="origin input" name="origin">
</div>

<input type="submit" value="Predict" class="submit-btn btn">
<p class="result" id="result">{{Prediction}} <span class="answer">{{mpg}}</span></p>
</form>
</div>
<p class="makers" >Team: Aakriti, Abhishek, Hoo Dharshana, Jayapradhap </p>
</body>
</html>
```

CSS code:

```
* {
box-sizing: border-box;
}

body {
font-family: 'Times New Roman', Times, serif;
background-image: url("bgp.jpg");
background-repeat: no-repeat;
background-size: cover;
background-position: center center;
padding: 0;
margin: 0;
}

header {
align-items: center;
justify-content: center;
width: 100%;
height: 40px;
display: flex;
}

h1 {
position: absolute;
text-align: center;
top: 2px;
color: #ffffff;
}

form {
width: max-content;
background-color: rgba(0, 0, 0, 0.5);
padding-left: 50px;
padding-right: 50px;
border-radius:20px;
padding-top: 10px;
```

```
}

.field-name {
padding: 10px 0;
margin: 0;
font-size: 20px;
font-weight: bolder;
}

.field {
padding: 2px 0;
}

.form-container {
font-family: 'Times New Roman', Times, serif;
display: flex;
align-items: center;
justify-content: center;
padding: 3vh;
min-height:max-content;
color: #ffffff;
}

.field input[type=number]{
width:200px;
background-color:#eeeef0;
font-size:15px;
padding:5px 10px;
color:black;
border-radius:20px;
border:none;
}

.submit-btn{
font-size:15px;
padding:10px 30px;
color:black;
```

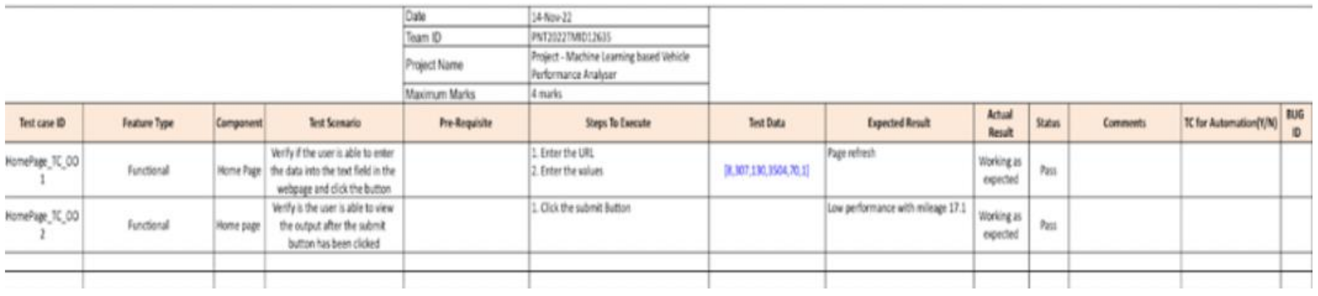
```
background-color:white;
border-radius:20px;
border:none;
display:block;
margin:20px auto;
cursor: pointer;
}

.submit-btn:hover{
color: white;
background-color: black;
border: none;
}

.result{
text-align:center;
font-size:20px;
color:white;
margin-top: 0;
text-decoration: underline;
}

.answer{
color:#ffffff;
font-weight:bolder;
}

.makers{
margin-top: 0%;
text-align: center;
color: white;
font-size:15px;
}
```

Python code:

```
from flask import Flask, render_template,request
import pickle
app=Flask(__name__)
model=pickle.load(open('RFregression.pkl','rb'))

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

@app.route('/model',methods=["GET","POST"])
def result():
no_of_clynder=request.form["no_of_cylinders"]
displacement=request.form["displacement"]
horsepower=request.form["horsepower"]
weight=request.form["weight"]
acceleration=request.form["acceleration"]
model_year=request.form["model_year"]
origin=request.form["origin"]

    t1=[[int(no_of_clynder),float(displacement),int(horsepower),int(weight),float(acceleration),int(m
    odel_year),int(origin)]]
output=model.predict(t1)
    return render_template("index.html", prediction= "The predicted MPG of the vehicle is ",
    mpg=str(output[0]))

if __name__ == "__main__":
app.run(debug=False)
```

8. TESTING

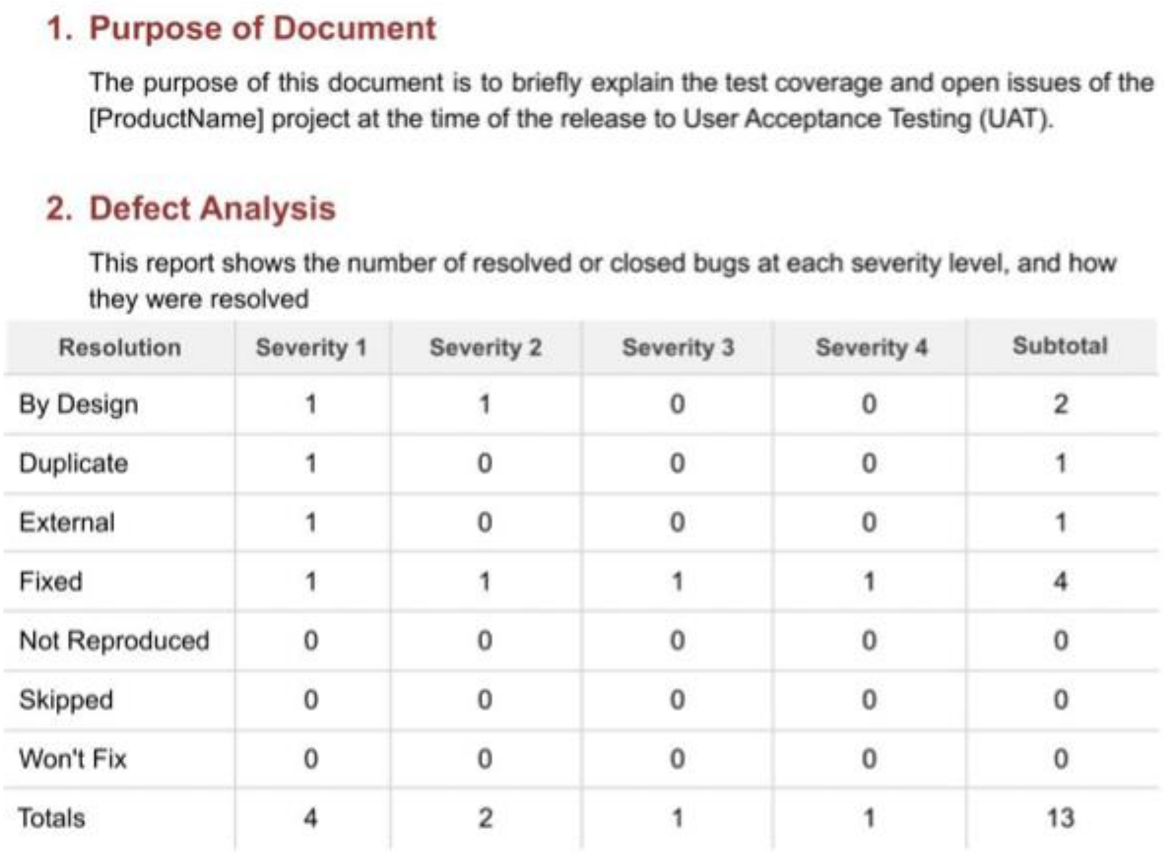
8.1 Test Cases



Date	14-Nov-22
Team ID	PNT2022TMID12635
Project Name	Project - Machine Learning based Vehicle Performance Analyser
Maximum Marks	4 marks

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID
HomePage_TC_OO1	Functional	Home Page	Verify if the user is able to enter the data into the text field in the webpage and click the button		1. Enter the URL 2. Enter the values	[8,307,130,3504,70,1]	Page refresh	Working as expected	Pass			
HomePage_TC_OO2	Functional	Home page	Verify is the user is able to view the output after the submit button has been clicked		1. Click the submit Button		Low performance with mileage 17.1	Working as expected	Pass			

8.2 User Acceptance Testing



1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the [ProductName] project at the time of the release to User Acceptance Testing (UAT).

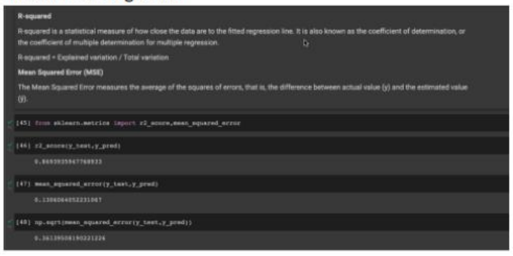
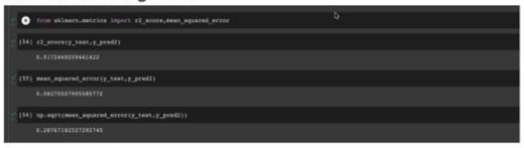
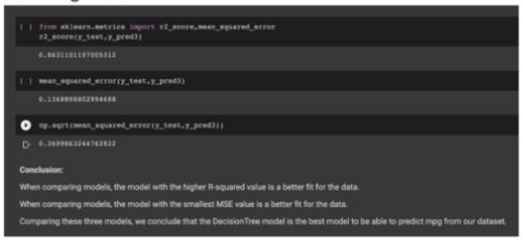
2. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	1	1	0	0	2
Duplicate	1	0	0	0	1
External	1	0	0	0	1
Fixed	1	1	1	1	4
Not Reproduced	0	0	0	0	0
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0
Totals	4	2	1	1	13

9. RESULTS

9.1 Performance Metrics

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Regression Model:** MAE - , MSE - , RMSE - , R2 score - **Classification Model:** Confusion Matrix - , Accuray Score- & Classification Report -	Decision tree regressor 
			Random forest regressor  Linear regression 

Decision tree regressor:

R-squared

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.

R-squared = Explained variation / Total variation

Mean Squared Error (MSE)

The Mean Squared Error measures the average of the squares of errors, that is, the difference between actual value (y) and the estimated value (ŷ).

```
[45] from sklearn.metrics import r2_score,mean_squared_error

[46] r2_score(y_test,y_pred)
     0.8693935947768933

[47] mean_squared_error(y_test,y_pred)
     0.1306064052231067

[48] np.sqrt(mean_squared_error(y_test,y_pred))
     0.36139508190221226
```

Random forest regressor:

```
from sklearn.metrics import r2_score,mean_squared_error

[34] r2_score(y_test,y_pred2)
     0.9172449209441422

[35] mean_squared_error(y_test,y_pred2)
     0.08275507905585772

[36] np.sqrt(mean_squared_error(y_test,y_pred2))
     0.28767182527292745
```

Linear regression:

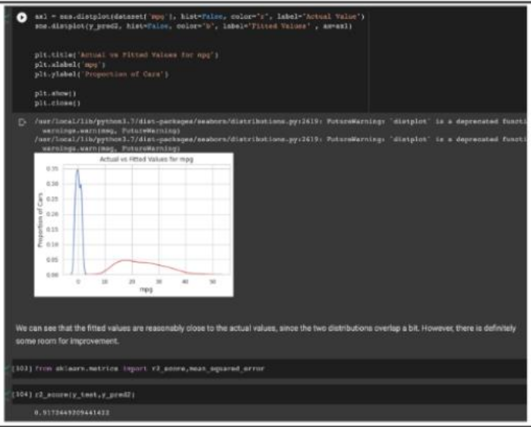
```
[ ] from sklearn.metrics import r2_score,mean_squared_error
    r2_score(y_test,y_pred3)
    0.8631101197005312

[ ] mean_squared_error(y_test,y_pred3)
    0.1368898802994688

np.sqrt(mean_squared_error(y_test,y_pred3))
    0.3699863244762822
```

Conclusion:

When comparing models, the model with the higher R-squared value is a better fit for the data.

When comparing models, the model with the smallest MSE value is a better fit for the data.

Comparing these three models, we conclude that the DecisionTree model is the best model to be able to predict mpg from our dataset.

Accuracy	Training Accuracy - 0.91724492094	

```
ax1 = sns.distplot(dataset['mpg'], hist=False, color="r", label='Actual Value')
sns.distplot(y_pred2, hist=False, color='b', label='Fitted Values' , ax=ax1)

plt.title('Actual vs Fitted Values for mpg')
plt.xlabel('mpg')
plt.ylabel('Proportion of Cars')

plt.show()
plt.close()

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated functi
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated functi
  warnings.warn(msg, FutureWarning)
```

We can see that the fitted values are reasonably close to the actual values, since the two distributions overlap a bit. However, there is definitely some room for improvement.

```
[103] from sklearn.metrics import r2_score,mean_squared_error

[104] r2_score(y_test,y_pred2)
      0.9172449209441422
```

10.ADVANTAGES & DISADVANTAGES

Advantages

• Using the Random Forest Algorithm in the model helps to perform both classifications as well as regression tasks.

• A random forest produces good predictions that can be easily understood.

• It can handle large datasets easily.

• Random Forest Algorithm provides a higher-level accuracy in predicting outcomes.

Disadvantages

• The main limitation of using random forest algorithm in the model is that a large number of trees can make the algorithm too slow and ineffective for real-time predictions.

• The random forest algorithm is quite slow to create predictions once it is trained.

11.CONCLUSION

The ability to estimate a car's performance level presents a big and fascinating challenge. Forecasting vehicle performance in order to improve particular vehicle behaviour was our main goal. Performance evaluation of the car considering its horsepower, cylinder count, acceleration, fuel type, and engine type, among other things. Based on the factors, like horsepower, cylinder count, fuel type, and engine type, the health of the car is forecasted. We analysed the components using a number of well-known machine learning approaches, like linear regression, decision trees, and random forests, in order to optimize the performance efficiency of the vehicle. The power, longevity, and range of automobile traction batteries are now the trends in automotive engineering. In this case, we additionally consider mileage performance. To answer this problem, we have built the models using a variety of methods and neural networks. We've then compared which algorithm is most accurate in forecasting car performance (Mileage). A front-end webpage was designed to help give the user an attractive front while they input the values required by the developed machine learning model. The IBM cloud platform was used to develop the model.

12.FUTURE SCOPE

The dataset used for this model is an old vehicle dataset, thus the model’s accuracy would drop when the details of vehicles released in recent times are given as input. Thus, in the future we propose to use the latest dataset set containing vehicle information to help train the model. We also plan to use other classification algorithms such as SVM and Decision Tress instead of Random Forest and measure if any accuracy gain occurs. Finally, we propose to scale the machine learning model to also analyse the performance of a larger range of vehicles.

13.APPENDIX

13.1 Source Code

https://github.com/IBM-EPBL/IBM-Project-26685-1660033317/tree/main/Application%20Building

13.2 GitHub & Project Demo Link

https://github.com/IBM-EPBL/IBM-Project-26685-1660033317

https://drive.google.com/file/d/1VY9ITMxCqJM_jz3lF01TcwQtFEY_OFs7/view?usp=share_link