TRIP BASED MODELING OF FUEL CONSUMPTION IN MORDERN FLEET USING MACHINE LEARNING

NALAIYA THIRAN PROJECT BASED LEARNING

on

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

A PROJECT REPORT

RIDDHIMA JANDLA	**19104136**
SREEDEV S NAIR	**19104155**
THEJUS S	**19104168**
VISHNU NARAYANAN S	**19104174**
VISHNU PRASAD M	**19104175**

BACHELOR OF ENGINEERING
IN

COMPUTER SCIENCE AND ENGINEERING

HINDUSTHAN COLLEGE OF ENGINEERING AND TECHOLOGY

Approved by AICTE, New Delhi, Accredited with ‘A’ Grade by NAAC

(An Autonomous Institution, Affiliated to Anna University, Chennai)

COIMBATORE – 641 032

November 2022

INTERNAL MENTOR

M AMUTHAVALLI

Department of Information Technology

Hindusthan College of Engineering and Technology,

Coimbatore-641 032

INDUSTRY MENTOR

TABLE OF CONTENTS

ABSTRACT

In August 2017 the government of India published final fuel-efficiency standards for commercial heavy-duty vehicles(HDV). The standards are the government's response to India's rapidly growing commercial vehicle sector. Driven by economic growth, India's diesel consumption has doubled in the past decade, increasing from 36.6 million metric tonnes in 2002 to 72.9 MMT in 2015. In India, HDVs account for a greater share of transportation petroleum end-use than in wealthier countries where passenger cars tend to dominate.

Predicting fuel consumption per trip based on dynamic on-road data can help the automotive industry to reduce the cost and time for on-road testing. Data modeling can easily help to diagnose the reason behind fuel consumption with a knowledge of input parameters

The ability to model and predict the fuel consumption is vital in enhancing fuel economy of vehicles and preventing fraudulent activities in fleet management. Fuel consumption of a vehicle depends on several internal factors such as distance, load vehicle characteristics, and driver behavior, as well as external factors such as road conditions, traffic, and weather. However, In this paper,(ANN) an artificial neural network was implemented to model fuel consumption in modern heavy-duty trucks for predicting the total and instantaneous fuel consumption of a trip based on very few key parameters, such as engine load (%),engine speed (rpm), and vehicle speed (km/h). The Instantaneous fuel consumption data can help to predict patterns in fuel consumption for optimized fleet operations. Hence, the challenge is to model or predict the fuel consumption only with the available data. Machine Learning (ML) is suitable in such analysis, as the model can be developed by learning the patterns in data. Also in this paper,we compare the ability of the ML techniques in predicting the fuel consumption of the Heavy Duty Vechicle. Moreover Predicting fuel consumption per trip based on dynamic on-road data can help the automotive industry to reduce the cost and time for on-road testing. Data modeling can easily help to diagnose the reason behind fuel consumption with a knowledge of input parameters.

In this paper, (ANN) an artificial neural network was implemented to model fuel consumption in modern heavy-duty trucks for predicting the total and instantaneous fuel consumption of a trip based on very few key parameters, such as engine load (%),engine speed (rpm), and vehicle speed (km/h). Instantaneous fuel consumption data can help to predict patterns in fuel consumption for optimized fleet operations.

INTRODUCTION

The fuel efficiency of heavy-duty trucks can be beneficial for the automotive and transportation industry, also for a country's economy and the global environment. The Reduction in the fuel consumption by just a few percent can significantly reduce costs for the transportation industry.
As per the All India study report submitted to PPAC, 70% of diesel and 99.6% petrol is consumed in the transport sector alone. Of the total diesel sale, the highest consumption of 28.48 is by cars, utility vehicles (UVs) and 3-wheelers. It was also revealed that private cars & UVs account for 13.15%.It was also revealed that in the States of Odisha, Bihar and Rajasthan, petrol consumption by two-wheelers exceeds 70%.
The impact of road infrastructure, traffic conditions, drivers' behavior, weather conditions, and the ambient temperature on fuel consumption were studied, and it was determined that fuel consumption can be reduced by 10% with eco-driving influences. with the help of machine learning techniques such as support vector machine (SVM), random forest (RF), and artificial neural networks (ANN) are widely applied to turn data into meaningful insights and solve complex problems.
These techniques have been applied to estimate emissions and fuel consumption in motor vehicles, trucks, ships, and aircraft. While the current approaches determine the fuel consumption of the vehicle, combining these techniques with data helps to identify parameters that may cause anomalies, such as malfunctions due to wear and tear of the engine, improper maintenance, engine failure, exhaust after-treatment system, and external factors like climate, traffic, road conditions, etc.

OBJECTIVE

By the end of this Project, you will:

Know fundamental concepts and can work on IBM Cognos Analytics.

- You'll be able to understand the problem to classify if it is a regression or a classification kind of problem.
- You will be able to know how to pre-process/clean the data using different data pre-processing techniques.
- Applying different algorithms according to the dataset
- You will be able to know how to find the accuracy of the model.
- You will be able to build web applications using the Flask framework.

3. IDEATION PHASE

3.1 LITERATURE SURVEY

1. Trip Based Modeling of Fuel Consumption in Modern Heavy-Duty Vehicles Using Artificial Intelligence

Author: Sasanka Katreddi and Arvind Thiruvengadam

https://www.mdpi.com/1413994

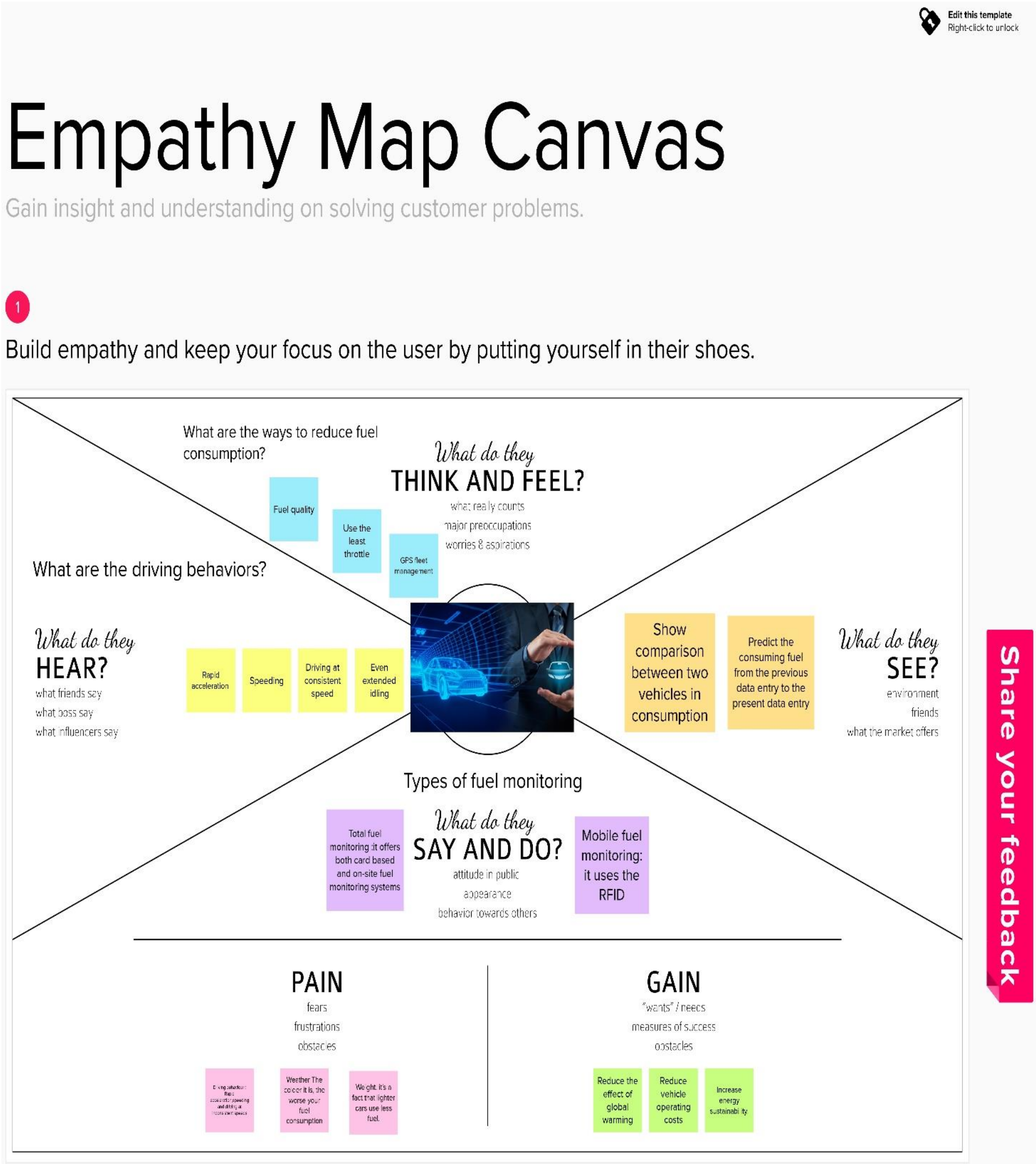
Heavy-duty trucks contribute approximately 20% of fuel consumption in the United States of America (USA). The fuel economy of heavy-duty vehicles (HDV) is affected by several real-world parameters like road parameters, driver behaviour, weather conditions, and vehicle parameters, etc. Although modern vehicles comply with emissions regulations, potential malfunction of the engine, regular wear and tear, or other factors could affect vehicle performance. Predicting fuel consumption per trip based on dynamic on-road data can help the automotive industry to reduce the cost and time for on-road testing. Data modelling can easily help to diagnose the reason behind fuel consumption with a knowledge of input parameters. In this paper, an artificial neural network (ANN) was implemented to model fuel consumption in modern heavy-duty trucks for predicting the total and instantaneous fuel consumption of a trip based on very few key parameters, such as engine load (%), engine speed (rpm), and vehicle speed (km/h). Instantaneous fuel consumption data can help to predict patterns in fuel consumption for optimized fleet operations. In this work, the data used for modelling was collected at a frequency of 1Hz during on-road testing of modern heavy-duty vehicles (HDV) at the West Virginia University Center for Alternative Fuels Engines and Emissions (WVU CAFEE) using the portable emissions monitoring system (PEMS). The performance of the artificial neural network was evaluated using mean absolute error (MAE) and root mean square error (RMSE). The model was further evaluated with data collected from a vehicle on-road trip. The study shows that artificial neural networks performed slightly better than other machine learning techniques such as linear regression (LR), and random forest (RF), with high R-squared (R^2) and lower root mean square error.

2. A Machine Learning Model for Average Fuel Consumption in Heavy Vehicles

Author: Rishikesh Mahesh Bagwe, Brent Hendrix, Alexander Schoen and Andy Byerly
https://www.researchgate.net/publication/333367045_A_Machine_Learning_Model_for_Average_Fuel_Consumption_in_Heavy_Vehicles

This paper advocates a data summarization approach based on distance rather than the traditional time period when developing individualized machine learning models for fuel consumption. This approach is used in conjunction with seven predictors derived from vehicle speed and road grade to produce a highly predictive neural network model for average fuel consumption in heavy vehicles. The proposed model can easily be developed and deployed for each individual vehicle in a fleet in order to optimize fuel consumption over the entire fleet. The predictors of the model are aggregated over fixed window sizes of distance travelled. Different window sizes are evaluated and the results show that a 1 km window is able to predict fuel consumption with a 0.91 coefficient of determination and mean absolute peak-to-peak percent error less than 4% for routes that include both city and highway duty cycle segments.

3.2 EMPATHY MAP



3.3 PROBLEM STATEMENT

1. Who does the Problem Affect?

Research confirms that vehicle emissions are responsible for 45% of the pollutants in the environment.

2. What is the issue?

   Rapid acceleration, speeding, driving at inconsistent speeds and even extended idling can increase your fuel consumption.

3. When does the issue occur?

Increasing your highway cruising speed from 55mph (90km/h) to 75mph (120km/h) can raise fuel consumption as much as 20%. It can improve your gas mileage 10 – 15% by driving at 55mph rather than 65mph(104km/h).

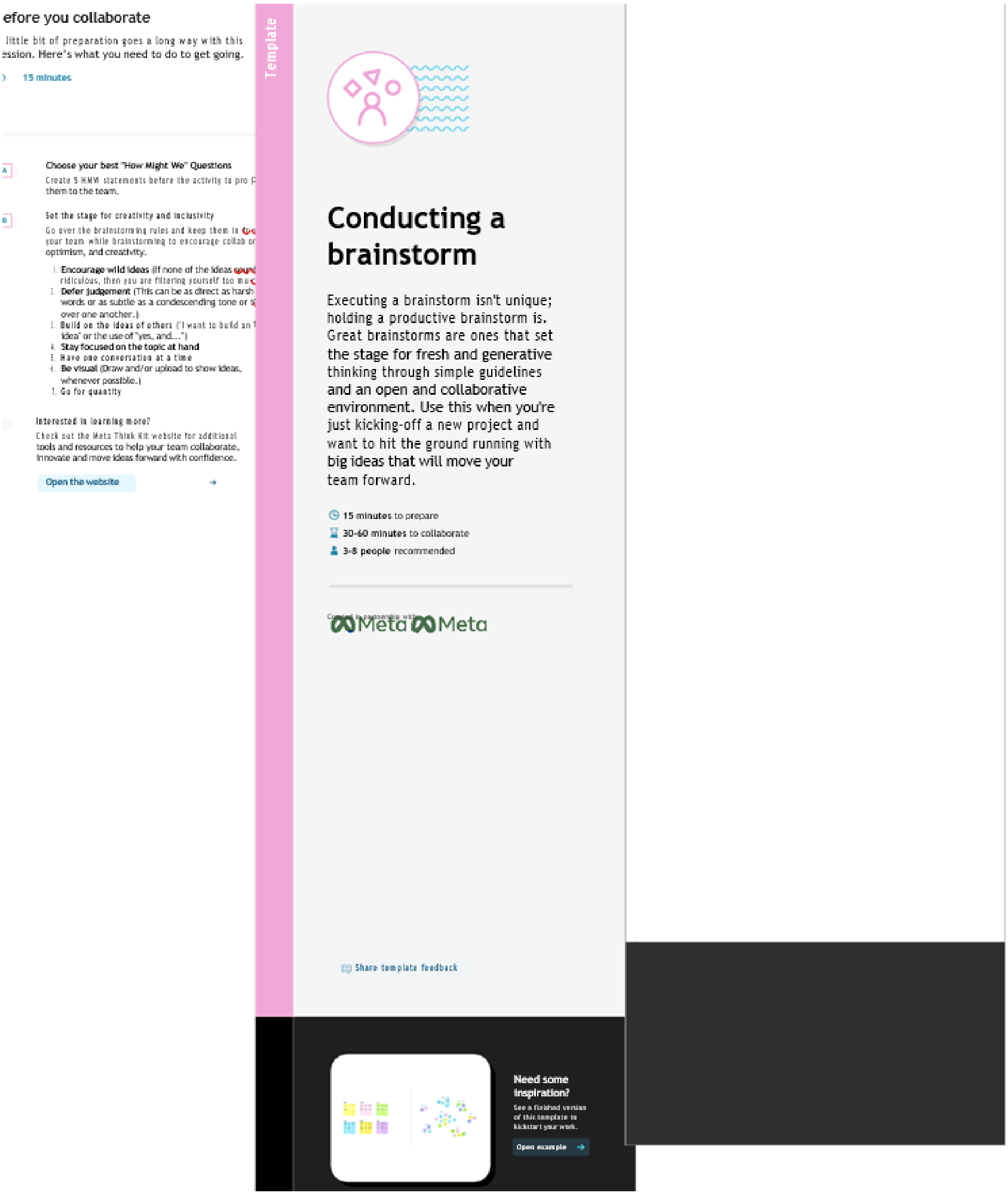
4. Where is the issue coming?

The issues commonly present in Oil, Oxygen sensors, air filters, spark plugs,and fuel injectors, Tires, Air Conditioning, Idling, Using the Wrong Gear, Aggressive Driving.

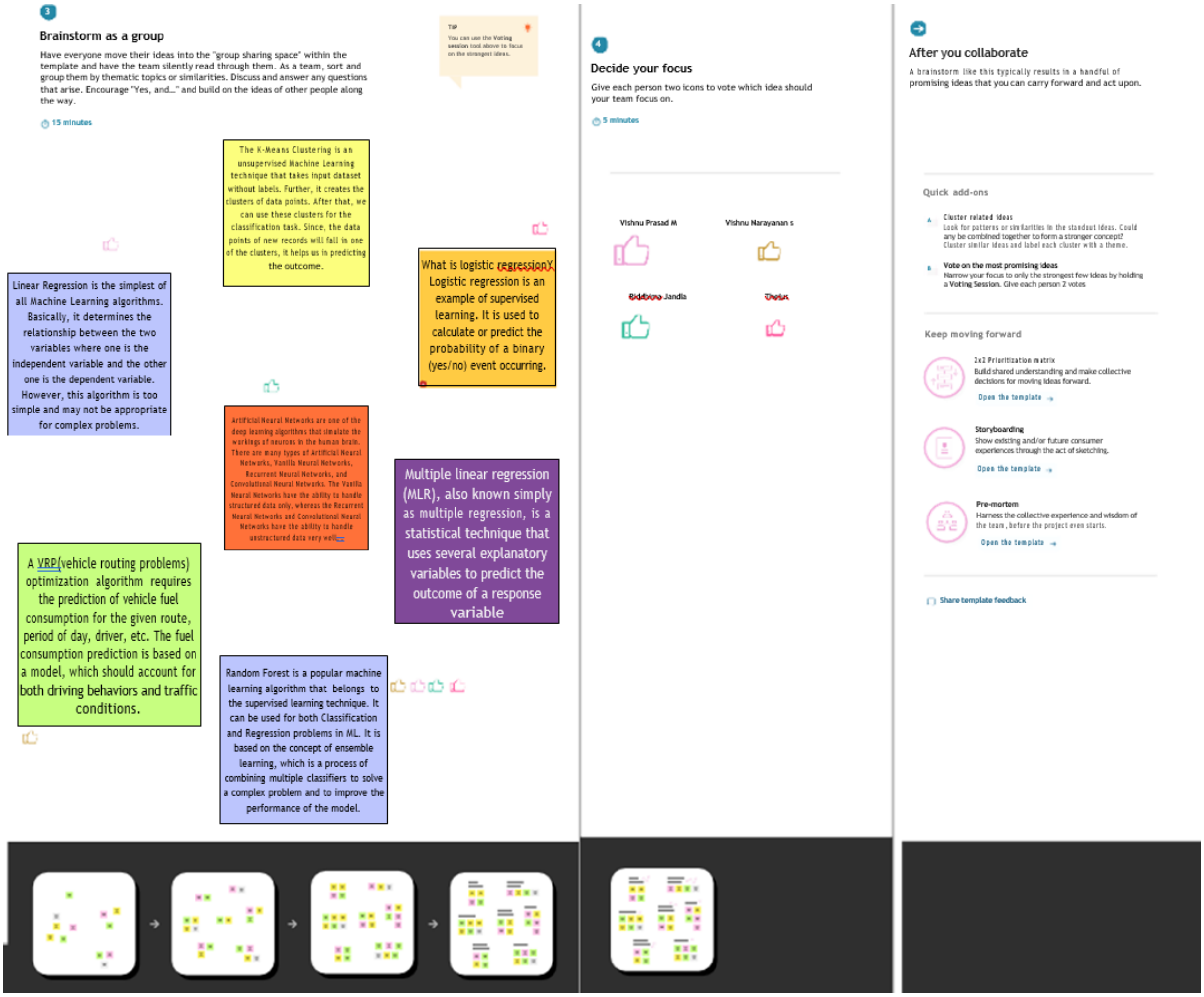
5. Why is it important that we fix the problem?

In addition to reducing the effects of global warming, reduced vehicle fuel consumption could arguably alleviate concerns regarding the future price and availability of fossil fuels. Reduced vehicle operating costs would be an obvious benefit resulting from reduced fuel consumption.

3.4 BRAIN STROMING AND IDEA




1
Choose your best "How Might We" Questions
Share the top 5 brainstorm questions that you created and let the group determine where to begin by selecting one question to move forward with based on what seems to be the most promising for idea generation in the areas you are trying to impact.
10 minutes
QUESTION
How might fleet managers predict fuel consumption accurately only using available data?
QUESTION
How does fuel consumption affect your fleet's running cost?
QUESTION
how to predict and manage fuel tax credit
QUESTION
How can we take measures to Rightsizing our Vehicle Fleet to Conserve Fuel
QUESTION
Why is important to model and predict the fuel consumption?
2
Brainstorm solo
Have each participant begin in the "solo brainstorm space" by silently brainstorming ideas and placing them into the template. This "silent-storming" avoids group-think and creates an inclusive environment for introverts and extroverts alike. Set a time limit. Encourage people to go for quantity.
10 minutes
EBM algorithm explains the relationship between the input factors and fuel consumption, quantifying the individual contribution of each one of them.
Ability to model and predict the fuel consumption is vital in enhancing fuel economy of vehicles and preventing fraudulent activities in fleet management.
Fleet fuel economy can be improved by eliminating extra weight, choosing efficient engine and vehicle parts, using effective fuel tracking technology
By predicting fuel consumption in fleet vehicles, can prove useful in planning trips as well as performing real-time predictions during driving can be predicted indirectly.
Commonly used models for these purposes are artificial neural networks (NNs), because they are universal approximators that can represent nonlinear characteristics of a complex system by using a nonlinear activation function
Based on many analysis, it can be concluded that the random forest technique produces a more accurate prediction compared to both the gradient boosting and neural networks.
Distance is not the only factor that affects fuel consumption. Multiple factors like speed, temperatures inside and outside, AC, and other weather conditions like rain or sun besides distance should be used to predict the consumption.
Distance is not the only factor that affects fuel consumption. Multiple factors like speed, temperatures inside and outside, AC, and other weather conditions like rain or sun besides distance should be used to predict the consu.
Vishnu Narayanan s
Create custom map zones (geofences) for improved off-road claims accuracy, and review and adjust fuel usage details if required. Maximise FTC claims via GPS Data Reports to enable you or your advisor to apportion fuel usage between activities Receive ongoing monthly FTC rebates at the optimal rates.
Fleet fuel economy is calculated using the harmonic mean, not a simple arithmetic mean. The harmonic mean is the reciprocal of the average of the reciprocals of the fuel economies of all of the vehicles in the fleet.
Transition to Smaller More Efficient Engines: Using smaller engines can help fleets meet operational needs without downgrading vehicle class. Some fleets choose to switch from 6-cylinder to 4-cylinder engines to help reduce fuel use and emissions.
A fleet rightsizing strategy should evaluate the business case of each vehicle to determine whether reassigning, replacing, or eliminating the vehicle would reduce fuel and maintenance costs without compromising fleet activities.
Convert to electric vehicles or alternative fuels. Fuel efficient vehicles require less gas to go a given distance. When we burn less gas, we cut global warming emissions and produce less pollution, while spending less on gas—a lot less.
Set an internal utilization threshold for your fleet's vehicles Determine the utilization of each vehicle Determine the number of over and under utilized vehicles Then Implement changes according to it.
Choose Lighter Vehicles: When purchasing new vehicles, look for opportunities to reduce vehicle weight. Lightweight materials, such as aluminum frames and smaller components can reduce rolling resistance and drag
Make Smart Vehicle Purchases. Fleet managers may decide to replace older vehicles with more fuel-efficient or alternative fuel vehicles. These purchasing strategies may help fleet managers make decisions that meet operational needs and conserve fuel

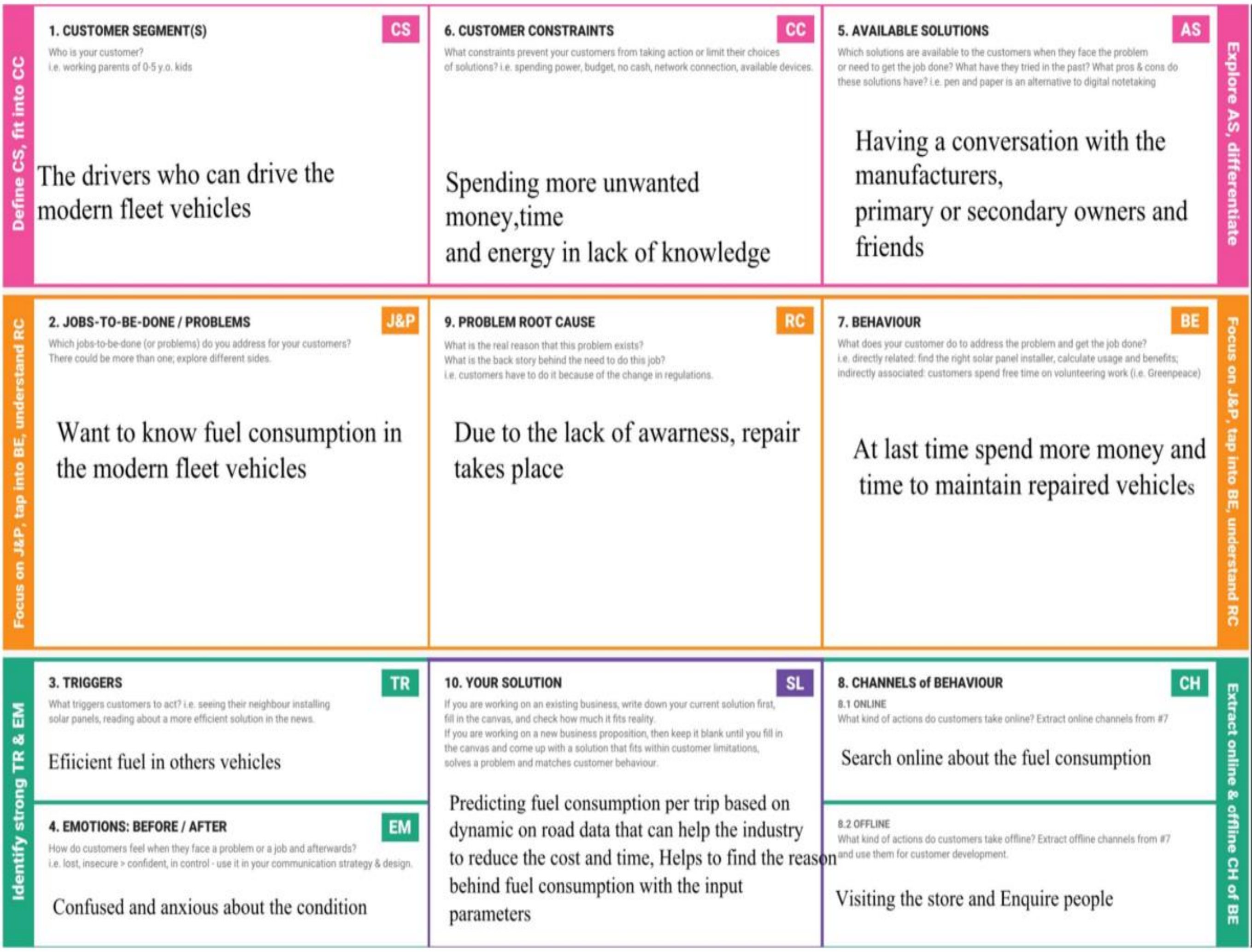

3
Brainstorm as a group
Have everyone move their ideas into the "group sharing space" within the template and have the team silently read through them. As a team, sort and group them by thematic topics or similarities. Discuss and answer any questions that arise. Encourage "Yes, and..." and build on the ideas of other people along the way.
15 minutes
TIP
You can use the Voting session tool above to focus on the strongest ideas.
The K-Means Clustering is an unsupervised Machine Learning technique that takes input dataset without labels. Further, it creates the clusters of data points. After that, we can use these clusters for the classification task. Since, the data points of new records will fall in one of the clusters, it helps us in predicting the outcome.
Linear Regression is the simplest of all Machine Learning algorithms. Basically, it determines the relationship between the two variables where one is the independent variable and the other one is the dependent variable. However, this algorithm is too simple and may not be appropriate for complex problems.
What is logistic regressionY Logistic regression is an example of supervised learning. It is used to calculate or predict the probability of a binary (yes/no) event occurring.
Artificial Neural Networks are one of the deep learning algorithms that simulate the workings of neurons in the human brain. There are many types of Artificial Neural Networks, Vanilla Neural Networks, Recurrent Neural Networks, and Convolutional Neural Networks. The Vanilla Neural Networks have the ability to handle structured data only, whereas the Recurrent Neural Networks and Convolutional Neural Networks have the ability to handle unstructured data very well.
Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable
A VRP(vehicle routing problems) optimization algorithm requires the prediction of vehicle fuel consumption for the given route, period of day, driver, etc. The fuel consumption prediction is based on a model, which should account for both driving behaviors and traffic conditions.
Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.
4
Decide your focus
Give each person two icons to vote which idea should your team focus on.
5 minutes
Vishnu Prasad M
Vishnu Narayanan s
Biddbiwa Jandia
Thelus
After you collaborate
A brainstorm like this typically results in a handful of promising ideas that you can carry forward and act upon.
Quick add-ons
Cluster related ideas
Look for patterns or similarities in the standout ideas. Could any be combined together to form a stronger concept? Cluster similar ideas and label each cluster with a theme.
Vote on the most promising ideas
Narrow your focus to only the strongest few ideas by holding a Voting Session. Give each person 2 votes
Keep moving forward
2x2 Prioritization matrix
Build shared understanding and make collective decisions for moving ideas forward.
Open the template
Storyboarding
Show existing and/or future consumer experiences through the act of sketching.
Open the template
Pre-mortem
Harness the collective experience and wisdom of the team, before the project even starts.
Open the template
Share template feedback

4. PROJECT DESIGN PHASE 1

4.1 PROPOSED SOLUTION

S .No	Parameter	Description
1.	Problem Statement (Problem to be solved)	Ability to model and predict the fuel consumption is vital in enhancing fuel economy of vehicles and preventing fraudulent activities in fleet management. Fuel consumption of a vehicle depends on several internal & external factors However, not all these factors may be measured or available for the fuel consumption analysis. The main aim of the project is to build Machine Learning algorithm to predict the fuel consumption of fleet vehicles based on the gas type . A web application is built which is integrated with ML model.
2.	Idea / Solution description	If you're driving an automatic car, make use of cruise control to keep your speed constant. And if you're driving a manual car, maintain a higher gear when appropriate. In each of these instances, your engines go through less revolutions per minute (RPM) and will reduce your fuel consumption.
3.	Novelty / Uniqueness	By using this project, the user can frequently monitoring the level of fuel consumption and he/she can take the necessary decision regarding the problematic situation.
4.	Social Impact / Customer Satisfaction	It does not require any expensive hardware for monitoring the fuel consumption level which makes the project efficiently to the customer.
5.	Business Model (Revenue Model)	By using this model the driver can avoid the unnecessary futuristic problems.
6.	Scalability of the Solution	By applying the machine learning algorithm, this solution provides a scalable solution to many number of customers at a time.

4.2 PROBLEM SOLUTION FIT



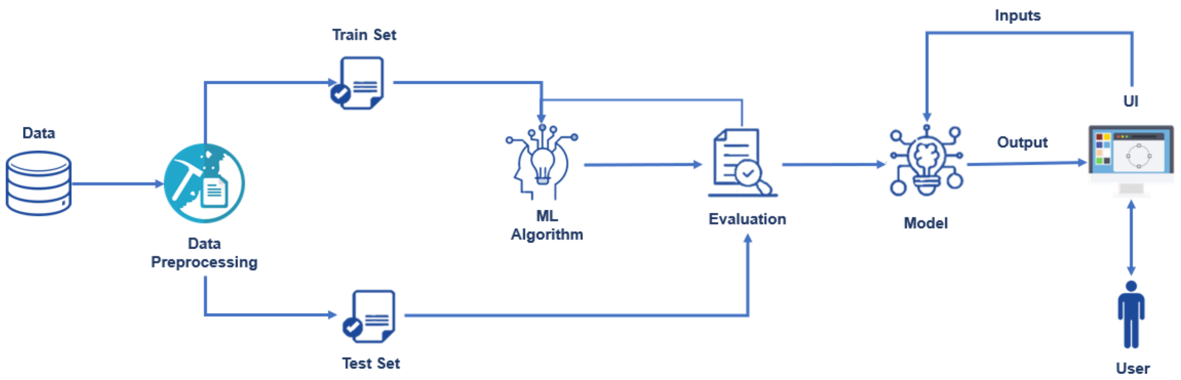
4.3 SOLUTION ARCHITECTURE

Solution Architecture:

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behaviour, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.

Example - Solution Architecture Diagram:



5.PROJECT DESIGN PHASE 2

5.1 CUSTOMER JOURNEY MAP

Customer Journey: A Day in the Life	**Morning**	**Afternoon**	**Night**
What is the customer thinking or feeling?	Condition of the vehicle.	Think about the fuel level after reaching some place.	Think about overall conditions of the vehicle.
What are the customer’s actions or main priorities?	Cover different sections of vehicles Such as fuel, engine, speed, etc.	Prefer a proper notification about the fuel level.	Perceive the future as better than the past.
What are the customer’s biggest pain points at this time?	Not able to find the necessary station if the fuel is empty.	Due to the heat, the fuel consumed too high, so it makes the user to drive properly .	Overwhelmed indication.
How does the customer interact with our product at this time?	Embracing the Day.	It prevents the user to waste the money.	Comforter.

5.2 SOLUTION REQUIREMENTS

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIn
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Transaction Processing	Payment through online Payment through offline
FR-4	Authentication	Through Email Physical Verification
FR-5	Reporting	Through App Through Email Through SMS

Non-functional Requirements:

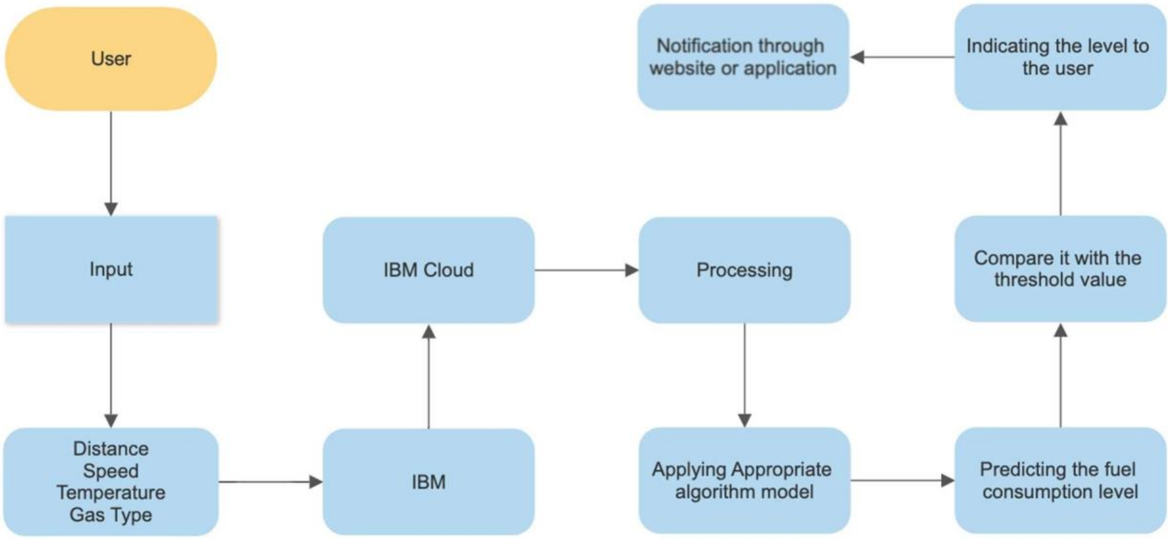
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	24/7
NFR-2	**Security**	Access permission for the particular system Information may only be changed by the system's data administrator
NFR-3	**Reliability**	Automatic restart and operation recovery
NFR-4	**Performance**	The front-page load time must be no more than 2 seconds for users that access the website using an LTE mobile connection.
NFR-5	**Availability**	New module deployment mustn't impact front page, product pages, and check out pages availability and mustn't take longer than one hour. The rest of the pages that may experience problems must display a notification with a timer showing when the system is going to be up again.
NFR-6	**Scalability**	The website/app attendance limit must be scalable enough to support 200,000 users at a time.

5.3 DATA FLOW DIAGRAM

Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD candepict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application byentering my email, password, and confirming my password.	I can access my account /dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation emailonce I have registered for the application	I can receive confirmationemail & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the applicationthrough Facebook	I can register & access thedashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the applicationthrough Gmail	I can access my account	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application byentering email & password	I can receive confirmationEmail and password through email	High	Sprint-1
	Dashboard	USN-6	As a user, I will able to see login and profiledetails	I can able to login through Gmail account	Medium	Sprint-3
Customer (Webuser)	Registration USN-	USN-7	As a user, I can register for the applicationby entering my name, phone number, email, password, and confirming my password.	I can access my googleaccount / dashboard	High	Sprint-1
Customer Care Executive	Chat Box	USN-8	As a customer care executive, I can ask about the inconvenience and the usabilityabout the app	I can access through IBM.Watson studio	Medium	Sprint-4
Administrator	Login	USN-9	As an administrator, I can help to loginand register the user through gmail	I can receive confirmation Gmail touser	High	Sprint-2
Fuel consumption predictor	IBM cloud	USN-10	As a tracker, I collect all the input ofvehicles from cloud	I can access the IBM Cloud	Medium	Sprint-1
Database Manager	IBM DB2	USN-11	As a database manager, I can collect all the user details and input for predicting the fuel consumption	I can access the IMBDb2	Medium	Sprint-3

5.4 TECHNOLOGY STACK

Technical Architecture:

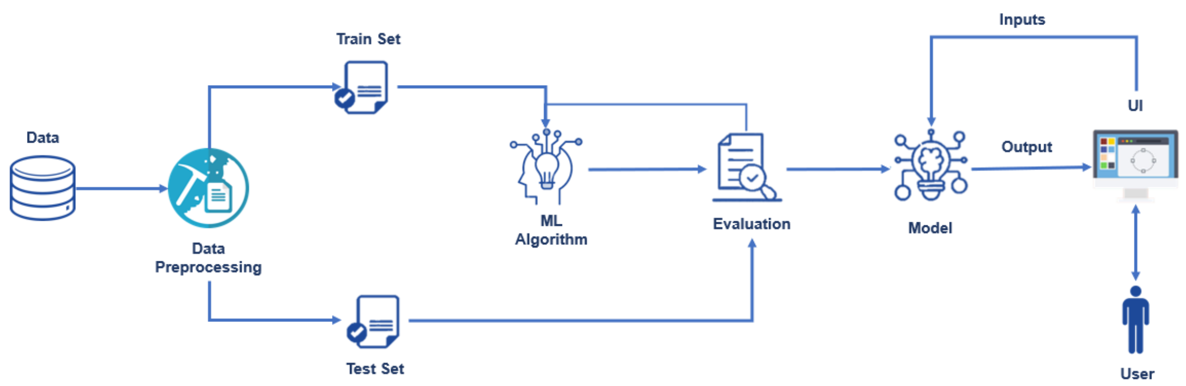


Table-1 : Components & Technologies:

Sl.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript / Angular Js / React Js etc.
2.	Application Logic-1	Logic for a process in the application	Java / Python
3.	Application Logic-2	Logic for a process in the application	IBM Watson STT service
4.	Application Logic-3	Logic for a process in the application	IBM Watson Assistant
5.	Database	Data Type, Configurations etc.	MySQL, NoSQL, etc.
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloudant etc.
7.	File Storage	File storage requirements	IBM Block Storage or Other Storage Service or Local Filesystem
8.	External API-1	Purpose of External API used in the application	IBM Weather API, etc.
9.	External API-2	Purpose of External API used in the application	Aadhar API, etc.
10.	Machine Learning Model	Purpose of Machine Learning Model	Object Recognition Model, etc.
11.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration :	Local, Cloud Foundry, Kubernetes, etc.

Table-2: Application Characteristics:

Sl.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	Technology of Opensource framework
2.	Security Implementations	List all the security / access controls implemented, use of firewalls etc.	e.g. SHA-256, Encryptions, IAM Controls, OWASP etc.
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	Technology used
4.	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)	Technology used
5.	Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.	Technology used

6. PROJECT PLANNING PHASE

6.1 PREPARE MILESTONE AND ACTIVITY LIST

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application byentering my username, email, password, organization name, and employee position.	4	High	Vishnu Prasad M Thejus
Sprint-1	Login	USN-2	As a user, I can log into the application by entering username/email & password	3	High	Vishnu Narayanan S Riddhima Jandla

Sprint-2	Dashboard	USN-3	As a user, I will be able to see the profiledetails and login details	2	Medium	Sreedev s Nair Vishnu Prasad M
Sprint-2	Entering input	USN-4	After entering the dashboard, I will give the input parameters for prediction	2	Medium	Thejus Riddhima Jandla
Sprint-2	Viewing output	USN-5	As a user, I will get the predicted output	6	High	Vishnu NarayananS Sreedevs Nair
Sprint-3	Report generation	USN-6	As a user, I will get a detailed report of output	5	Medium	Vishnu Prasad M Riddhima Jandla
Sprint-4	Instructions	USN-7	As a user, I can go through the instructions in the site to carry out the prediction process.	4	Low	Vishnu Narayanan S Thejus
Sprint-4	Portal Configuration	USN-8	As a developer, I can configure the portal foreasy use by customer	5	Medium	Vishnu Prasad M Sreedev s Nair

Project Tracker, Velocity & Burndown Chart:

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	7	6 Days	24 Oct 2022	29 Oct 2022		
Sprint-2	10	10 Days	31 Oct 2022	9 Nov 2022		
Sprint-3	5	2 Days	11 Nov 2022	12 Nov 2022		
Sprint-4	9	3 Days	14 Nov 2022	16 Nov 2022		

Velocity:

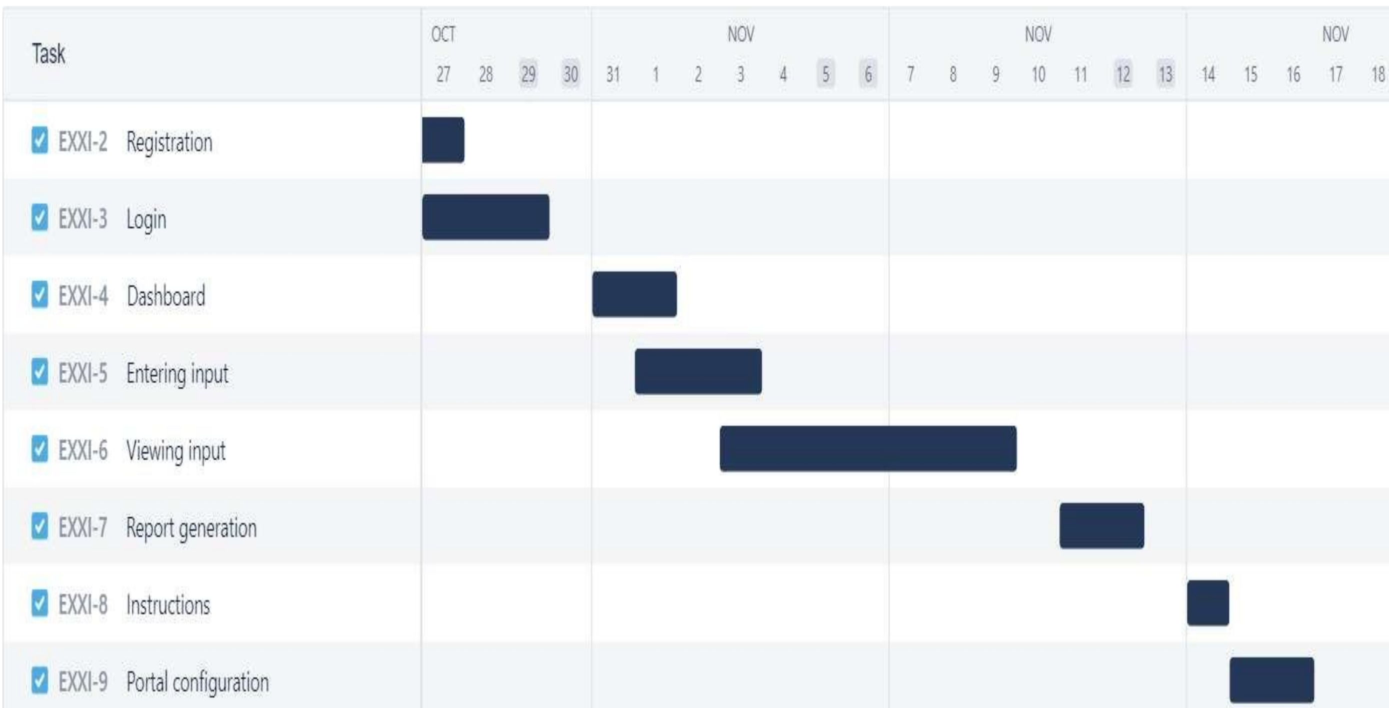
Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprit). Let's calculate the team's average velocity (AV) per iteration unit (story per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

Sprint	Average velocity
Sprint 1	1.17
Sprint 2	1
Sprint 3	2.5
Sprint 4	3

Burndown Chart:

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



6.2 Sprint Delivery Plan

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring the, technical papers, research publications etc.	3 SEPTEMBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	10 SEPTEMBER 2022
Ideation	List the by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	10 SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	2 OCTOBER 2022
Problem Solution Fit	Prepare problem - solution Fit document.	29 SEPTEMBER 2022
Solution Architecture	Prepare solution Architecture document	20 OCTOBER 2022
Customer Journey	Prepare the customer journey maps to understand the user interactions & experiences with the application	8 OCTOBER 2022
Data Flow Diagrams	Draw the data flow Diagrams and submit for review.	20 OCTOBER 2022
Technology Architecture Prepare Milestone & Activity List	Architecture diagram.	20 OCTOBER 2022
Project Development - Delivery of Sprint- 1, 2, 3 & 4	Develop & submit the developed code by testing it.	2 NOVEMBER 2022

7. PROJECT DEVELOPMENT PHASE

7.1 CODING & SOLUTIONING -SPRINT 1

In [2]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

In [4]:
```
df=pd.read_excel('measurements2.xlsx')
```

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

Out[6]:
```
(388, 12)
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 388 entries, 0 to 387
Data columns (total 12 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   distance       388 non-null    float64
 1   consume        388 non-null    float64
 2   speed          388 non-null    int64
 3   temp_inside    376 non-null    float64
 4   temp_outside   388 non-null    int64
 5   specials       93 non-null     object
 6   gas_type       388 non-null    object
 7   AC             388 non-null    int64
 8   rain           388 non-null    int64
 9   sun            388 non-null    int64
 10  refill liters  13 non-null     float64
 11  refill gas     13 non-null     object
dtypes: float64(4), int64(5), object(3)
memory usage: 36.5+ KB
```

In [8]:
```
df.describe()
```

Out[8]:

	distance	consume	speed	temp_inside	temp_outside	AC	rain	sun	refill liters
count	388.000000	388.000000	388.000000	376.000000	388.000000	388.000000	388.000000	388.000000	13.000000
mean	19.652835	4.912371	41.927835	21.929521	11.358247	0.077320	0.123711	0.082474	37.115385
std	22.667837	1.033172	13.598524	1.010455	6.991542	0.267443	0.329677	0.275441	8.587282
min	1.300000	3.300000	14.000000	19.000000	-5.000000	0.000000	0.000000	0.000000	10.000000
25%	11.800000	4.300000	32.750000	21.500000	7.000000	0.000000	0.000000	0.000000	37.600000
50%	14.600000	4.700000	40.500000	22.000000	10.000000	0.000000	0.000000	0.000000	38.000000
75%	19.000000	5.300000	50.000000	22.500000	16.000000	0.000000	0.000000	0.000000	39.000000

In [9]:
```
#Check Null Values
```

In [10]:
```
df.isnull().sum()
```

Out[10]:
```
distance          0
consume           0
speed             0
temp_inside      12
temp_outside      0
specials        295
gas_type          0
AC                0
rain              0
sun               0
refill liters   375
refill gas      375
dtype: int64
```

In [11]:
```
#Remove the null values
```

In [12]:
```
df.drop(['specials','refill liters','refill gas'],axis=1,inplace=True)
```

In [13]:
```
df.head(2)
```

Out[13]:

	distance	consume	speed	temp_inside	temp_outside	gas_type	AC	rain	sun
0	28.0	5.0	26	21.5	12	E10	0	0	0
1	12.0	4.2	30	21.5	13	E10	0	0	0

In [14]:
```
#Handling the null values
```

In [15]:
```
mn = df.temp_inside.mean()
mn
```

Out[15]:
```
21.929521276595743
```

In [16]:
```
med = df.temp_inside.median()
med
```

Out[16]:
```
22.0
```

In [17]:
```
df['temp_inside']=df.temp_inside.fillna(mn)
```

In [18]:
```
df.isnull().sum()
```

Out[18]:
```
distance        0
consume         0
speed           0
temp_inside     0
```

In [20]:
```
#Descriptive Statistics
```

In [21]:
```
df[['distance','consume','speed','temp_inside',
   'temp_outside']].mean()
```

Out[21]:
```
distance        19.652835
consume          4.912371
speed           41.927835
temp_inside     21.929521
temp_outside    11.358247
dtype: float64
```

In [22]:
```
df[['distance','consume','speed','temp_inside',
   'temp_outside']].median()
```

Out[22]:
```
distance        14.6
consume          4.7
speed           40.5
temp_inside     22.0
temp_outside    10.0
dtype: float64
```

In [23]:
```
df[['gas_type','AC','rain','sun']].mode()
```

Out[23]:

	gas_type	AC	rain	sun
0	SP98	0	0	0

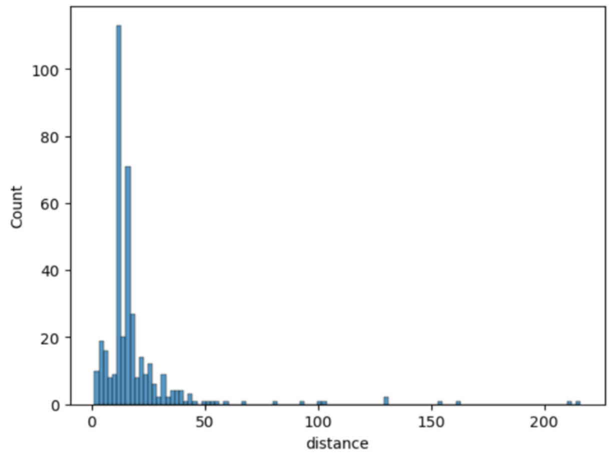
In [24]:
```
df.describe()
```

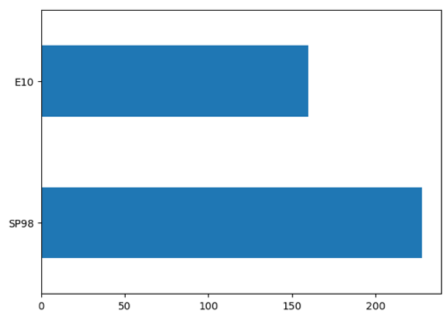
Out[24]:

	distance	consume	speed	temp_inside	temp_outside	AC	rain	sun
count	388.000000	388.000000	388.000000	388.000000	388.000000	388.000000	388.000000	388.000000
mean	19.652835	4.912371	41.927835	21.929521	11.358247	0.077320	0.123711	0.082474
std	22.667837	1.033172	13.598524	0.994666	6.991542	0.267443	0.329677	0.275441
min	1.300000	3.300000	14.000000	19.000000	-5.000000	0.000000	0.000000	0.000000
25%	11.800000	4.300000	32.750000	21.500000	7.000000	0.000000	0.000000	0.000000
50%	14.600000	4.700000	40.500000	22.000000	10.000000	0.000000	0.000000	0.000000
75%	19.000000	5.300000	50.000000	22.500000	16.000000	0.000000	0.000000	0.000000
max	216.100000	12.200000	90.000000	25.500000	31.000000	1.000000	1.000000	1.000000

```
In [26]: sns.histplot(df.distance)
```

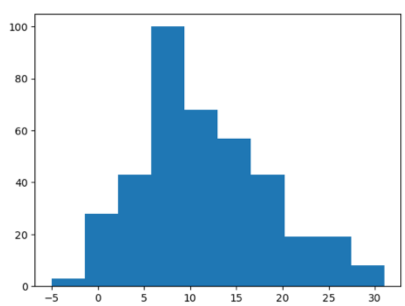
```
Out[26]: <AxesSubplot:xlabel='distance', ylabel='Count'>
```





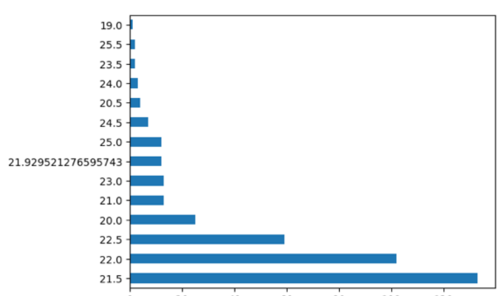

```
In [37]: plt.hist(df.temp_outside)
```

```
Out[37]: (array([  3.,  28.,  43., 100.,  68.,  57.,  43.,  19.,  19.,   8.]),
          array([-5. , -1.4,  2.2,  5.8,  9.4, 13. , 16.6, 20.2, 23.8, 27.4, 31. ]),
          <BarContainer object of 10 artists>)
```

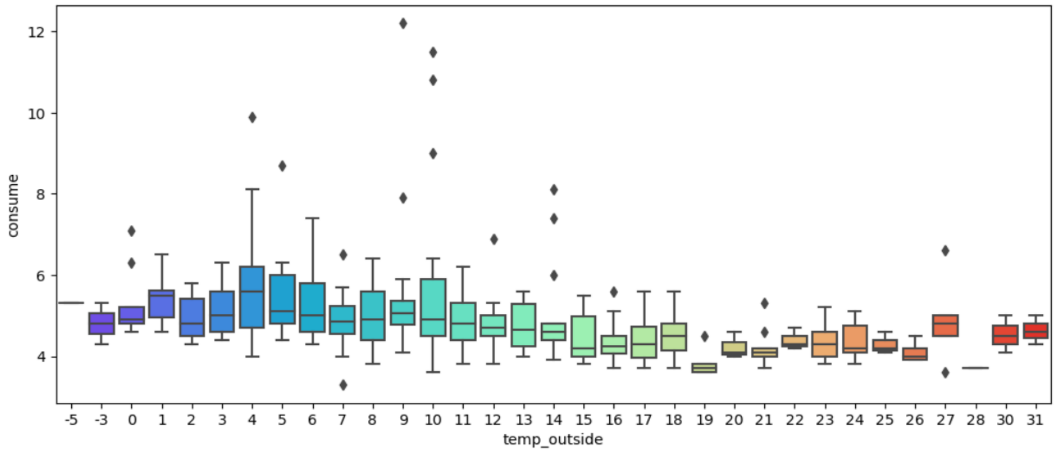


```
In [38]: df.temp_inside.value_counts().plot(kind='barh')
```

```
Out[38]: <AxesSubplot:>
```



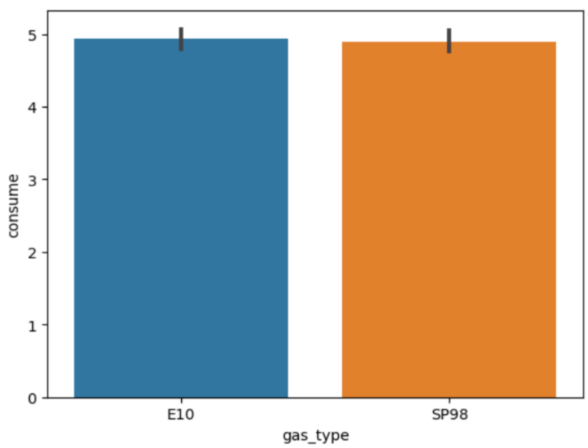
In [43]:
```
plt.figure(figsize=(12,5))
sns.boxplot(x='temp_outside',y='consume',data=df,palette='rainbow')
```
Out[43]: <AxesSubplot:xlabel='temp_outside', ylabel='consume'>



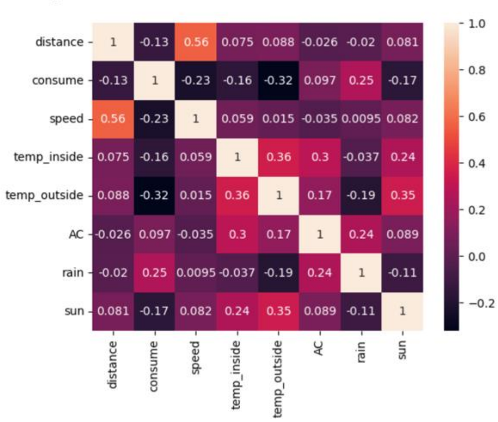
In [44]:
```
df.head(2)
```
Out[44]:

	distance	consume	speed	temp_inside	temp_outside	gas_type	AC	rain	sun
0	28.0	5.0	26	21.5	12	E10	0	0	0
1	12.0	4.2	30	21.5	13	E10	0	0	0

In [45]:
```
sns.barplot(x='gas_type',y='consume',data=df)
```
Out[45]: <AxesSubplot:xlabel='gas_type', ylabel='consume'>



In [49]:
```
#Multivariate Analysis
```
In [50]:
```
sns.heatmap(df.corr(),annot=True)
```
Out[50]: <AxesSubplot:>



7.2 CODING & SOLUTIONING -SPRINT 2

```
In [53]: from sklearn.preprocessing import LabelEncoder

In [54]: le=LabelEncoder()

In [55]: df['gas_types']=le.fit_transform(df.gas_type)

In [56]: df.drop('gas_type',axis=1,inplace=True)

In [57]: df.head(2)
Out[57]:
```

	distance	consume	speed	temp_inside	temp_outside	AC	rain	sun	gas_types
0	28.0	5.0	26	21.5	12	0	0	0	0
1	12.0	4.2	30	21.5	13	0	0	0	0

```
In [58]: x=df.drop(['consume'],axis=1)
         y=df.consume

In [59]: from sklearn.model_selection import train_test_split

In [60]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,
                                                        random_state=42)

In [61]: from sklearn.linear_model import LinearRegression

In [62]: linear_reg=LinearRegression()

In [63]: linear_reg.fit(x_train,y_train)
Out[63]: LinearRegression()

In [64]: y_pred=linear_reg.predict(x_test)

In [65]: from sklearn.metrics import mean_absolute_error,mean_squared_error

In [66]: mean_absolute_error(y_test,y_pred)
Out[66]: 0.6689496242764842

In [67]: mean_squared_error(y_test,y_pred)
Out[67]: 0.7476946880297469

In [68]: np.sqrt(mean_squared_error(y_test,y_pred))
Out[68]: 0.864693406954018
```

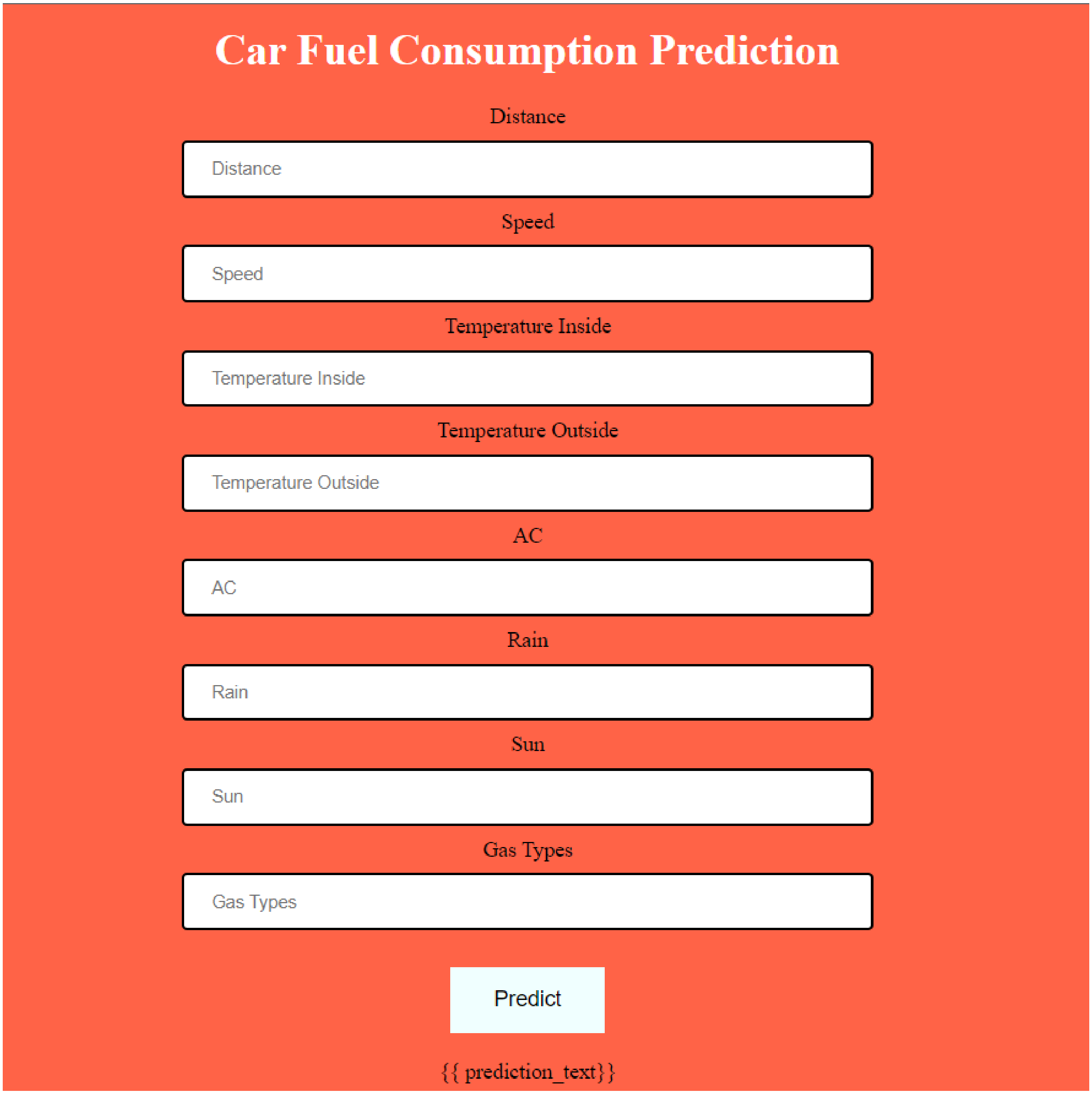
8. TESTING

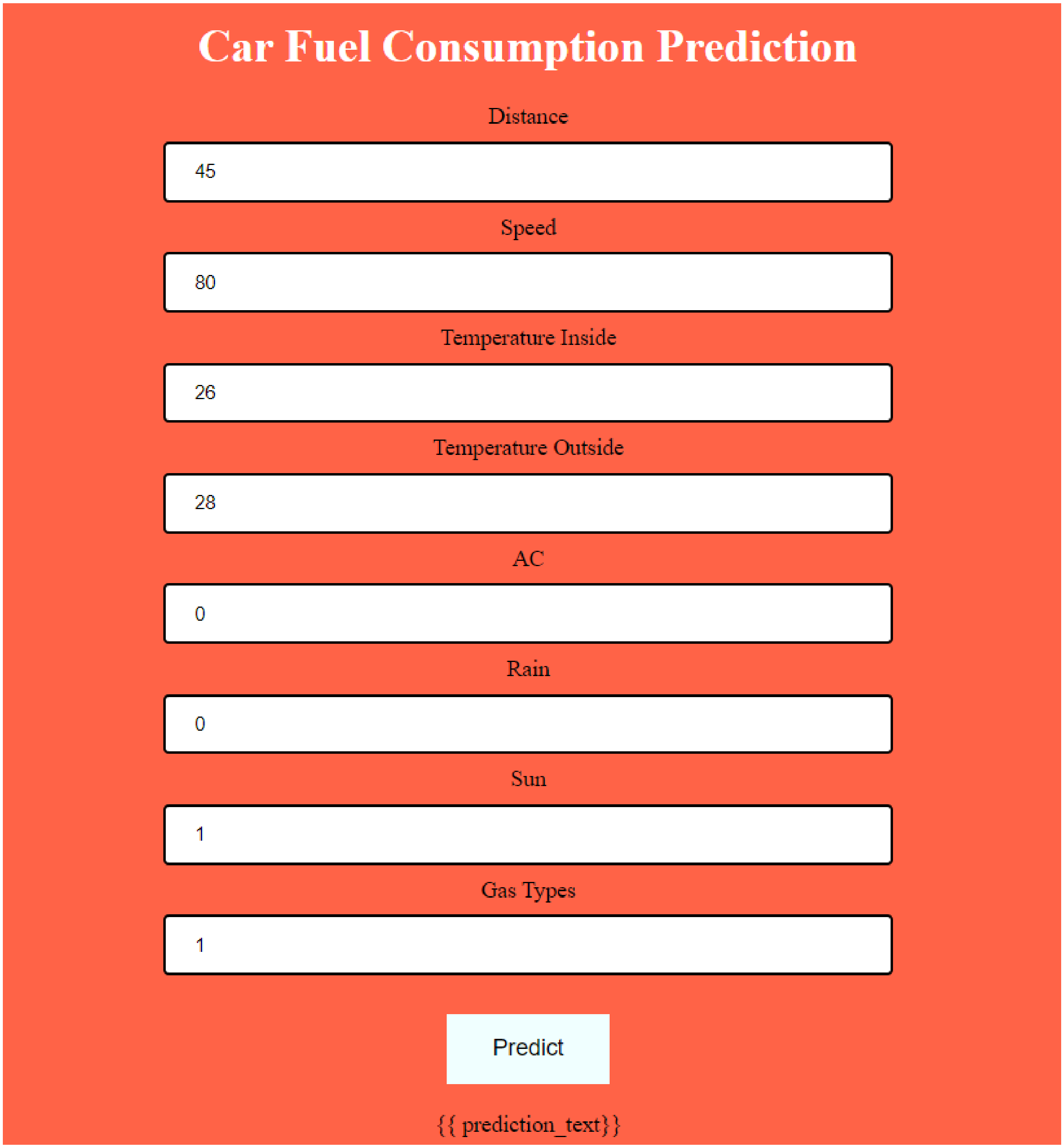
8.1 TEST CASES

Prediction

```
In [70]: linear_reg.predict([[16,39,24,18,0,0,0,1]])
Out[70]: array([4.3111517])

In [71]: linear_reg.predict([[18,49,21,20,0,2,0,0]])
Out[71]: array([5.7988122])

In [72]: linear_reg.predict([[16,62,22,16,1,2,1,1]])
Out[72]: array([5.74573626])

In [73]: linear_reg.predict([[30,26,18,26,0,1,1,0]])
Out[73]: array([5.95978692])
```

8.2 USER ACCEPTANCE TESTING




Car Fuel Consumption Prediction
Distance
45
Speed
80
Temperature Inside
26
Temperature Outside
28
AC
0
Rain
0
Sun
1
Gas Types
1
Predict
{{ prediction_text}}

9. RESULTS

9.1 PERFORMANCE MATRIX

```
In [62]: from sklearn.metrics import mean_absolute_error,mean_squared_error

In [63]: mean_absolute_error(y_test,y_pred)
Out[63]: 0.6689496242764842

In [64]: mean_squared_error(y_test,y_pred)
Out[64]: 0.7476946880297469

In [68]: np.sqrt(mean_squared_error(y_test,y_pred))
Out[68]: 0.864693406954018
```

10. ADVANTAGES & DISADVANTAGES

ADVANTAGES:

If you want to minimize the impact that your journeys have on the environment, fuel efficiency is an important consideration. Models with low fuel consumption use less petrol or diesel, both of which come from crude oil which is a finite resource. You should also look at CO2 emissions, as higher emissions will have a greater impact on environmental issues such as climate change.

Driving a fuel-efficient vehicle reduces the running costs you will have to pay throughout the lifetime of your vehicle. There are many different savings associated with fuel efficiency, and together they add up to a significant amount of money.

If you are looking for the best option for vehicles in your fleet, you will need to consider Whole Life Costs in order to work out which model is the most cost-effective in the long run. The amount you spend on fuel and Vehicle Excise Duty is a large contributory factor to the Whole Life Cost, so choosing a fuel-efficient car is a sound business decision.

DISADVANTAGES:

From an environmental perspective, fuel consumption results in the production of vehicle emissions which can be classified into air pollutants (which affect health) and greenhouse gases (which affect the environment). Fuel consumption also depletes stocks of non-renewable fossil fuels.

The global gasoline and diesel fuel vehicle fleets impose substantial impacts on air quality, human health, and climate changes.

Amount of money spend on fuel for vehicles in normal house hold is increasing day by day so it increase annual expenditure.

11. APPENDIX

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
df=pd.read_excel('measurements2.xlsx')
print(df.shape)
df.info()
```

```
df.describe()
df.isnull().sum()
df.drop(['specials','refill liters','refill gas'],axis=1,inplace=True)
mn = df.temp_inside.mean()
med = df.temp_inside.median()
df['temp_inside']=df.temp_inside.fillna(mn)
df.isnull().sum()
df[['distance','consume','speed','temp_inside', 'temp_outside']].mean()
df[['distance','consume','speed','temp_inside','temp_outside']].median()
df[['gas_type','AC','rain','sun']].mode()
df.describe()
sns.histplot(df.distance)
sns.kdeplot(df.distance,shade=True)
sns.histplot(df.speed)
sns.kdeplot(df.speed,shade=True)
sns.histplot(df.temp_inside)
sns.kdeplot(df.temp_inside,shade=True)
sns.histplot(df.temp_outside)
sns.kdeplot(df.temp_outside,shade=True)
plt.hist(df.gas_type)
plt.figure(figsize=(7,5))
df.gas_type.value_counts().plot(kind='barh')
plt.hist(df.temp_outside)
df.temp_inside.value_counts().plot(kind='barh')
plt.hist(df.temp_inside)
sns.barplot(x='gas_type',y='consume',data=df)
plt.figure(figsize=(12,5))
sns.boxplot(x='temp_outside',y='consume',data=df,palette='rainbow')
```

```
sns.barplot(x='gas_type',y='consume',data=df)
sns.barplot(x='AC',y='consume',data=df)
sns.barplot(x='rain',y='consume',data=df)
sns.barplot(x='sun',y='consume',data=df)
sns.heatmap(df.corr(),annot=True)
sns.pairplot(df)
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df['gas_types']=le.fit_transform(df.gas_type)
df.drop('gas_type',axis=1,inplace=True)
x=df.drop(['consume'],axis=1)
y=df.consume
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=4)
from sklearn.linear_model import LinearRegression
linear_reg=LinearRegression()
linear_reg.fit(x_train,y_train)
y_pred=linear_reg.predict(x_test)
from sklearn.metrics import mean_absolute_error,mean_squared_error
mean_absolute_error(y_test,y_pred)
mean_squared_error(y_test,y_pred)
np.sqrt(mean_squared_error(y_test,y_pred))
import pickle
pickle.dump(linear_reg,open('model.pkl','wb'))
import joblib
joblib.dump(linear_reg,'model.save')
```

12. CONCLUSION

In conclusion, the study demonstrates the modelling of fuel consumption in modern fleet vehicles using Machine learning. An attempt was made to develop a model using very few parameters collected under different conditions. Data from heavy vehicles with the same make and model, driven by different persons on various routes under different external conditions, were used. The data modelling can help to identify the trend in instantaneous fuel consumption and to calculate the total fuel consumed by the vehicles, which can further help in diagnosing vehicle performance in the case of abnormalities.