CAR RESALE VALUE PREDICTION

USING CLOUD

A Project report submitted in partial fulfilment of 7th semester in degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

Team ID: PNT2022TMID41109

DHANESHKUMAR K	612719104301
DHARMADURAI D	612719104302
NAVEEN K	612719104303
GOWTHAM M	612719104024

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

THE KAVERY ENGINEERING COLLEGE,SALEM

(ANNA UNIVERSITY)

NOV-2022

THE KAVERY ENGINEERING COLLEGE, SELAM
(ANNA UNIVERSITY)

BONAFIDE CERTIFICATE

Certified that this project report "**CAR RESALE VALUE PREDICTION"** is the bonafide record work done by **Mr DHANESHKUMAR K**(612719104301), **Mr DHARMADURAI D**(612719104302), **Mr NAVEEN K**(612719104303), **Mr GOWTHAM M**(6127104024) for **IBM-NALAIYATHIRAN** in **VII** semester of **B.E.,** degree course **in Computer Science and Engineering** branch during the academic year of 2022 - 2023.

Staff-In charge
Ms SUDHA G

Evaluator
Ms VASUKI S

Head of the Depart Department
Mr. M. BALAMURUGAN

ACKNOWLEDGEMENT

We are highly grateful to thank our Project coordinator MS SUDHA .G and our Project Evaluator MS VASUKI S Department of Computer Science and Engineering, The Kavery Engineering College, Salem for the coordinating us throughout this Project.

We are very much indebted to thank all the faculty members of Department of Computer Science and Engineering in our Institute, for their excellent moral support and suggestions to complete our Project work successfully.

DHANESHKUMAR K (612719104301)
DHARMADURAI D (612719104302)
NAVEEN K (612719104303)
GOWTHAM M (612719104024)

ABSTRACT

With difficult economic conditions, it is likely that sales of secondhand imported (reconditioned) cars and used cars will increase. In many developed countries, it is common to lease a car rather than buying it outright. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, it is of commercial interest to sellers/financers to be able to predict the salvage value (residual value) of cars with accuracy.

In order to predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale value of the vehicle. We will be using various regression algorithms and algorithm with the best accuracy will be taken as a solution, then it will be integrated to the web-based application where the user is notified with the status of his product

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that

Project Report Format

1. **INTRODUCTION**
   - 1.1 Project Overview
   - 1.2 Purpose
2. **LITERATURE SURVEY**
   - 2.1 Existing problem
   - 2.2 References
   - 2.3 Problem Statement Definition
3. **IDEATION & PROPOSED SOLUTION**
   - 3.1 Empathy Map Canvas
   - 3.2 Ideation & Brainstorming
   - 3.3 Proposed Solution
   - 3.4 Problem Solution fit
4. **REQUIREMENT ANALYSIS**
   - 4.1 Functional requirement
   - 4.2 Non-Functional requirements
5. **PROJECT DESIGN**
   - 5.1 Data Flow Diagrams
   - 5.2 Solution & Technical Architecture
   - 5.3 User Stories
6. **PROJECT PLANNING & SCHEDULING**
   - 6.1 Sprint Planning & Estimation
   - 6.2 Sprint Delivery Schedule
   - 6.3 Reports from JIRA
7. **CODING & SOLUTIONING (Explain the features added in the project along with code)**
   - 7.1 Feature 1
   - 7.2 Feature 2
   - 7.3 Database Schema (if Applicable)
8. **TESTING**
   - 8.1 Test Cases
   - 8.2 User Acceptance Testing
9. **RESULTS**
   - 9.1 Performance Metrics
10. **ADVANTAGES & DISADVANTAGES**
11. **CONCLUSION**
12. **FUTURE SCOPE**
13. **APPENDIX**

    Source Code

GitHub & Project Demo Link

CAR RESALE VALUE PREDICTION

1. INTRODUCTION:

In this project we have used different algorithms with different techniques for developing Car resale value prediction systems considering different features of the car. In a nutshell, car resale value prediction helps the user to predict the resale value of the car depending upon various features like kilometers driven, fuel type, etc.

1.1 Project overview:

Car resale value prediction system is made with the purpose of predicting the correct valuation of used cars that helps users to sell the car remotely with perfect valuation and without human intervention in the process to eliminate biased valuation. Due to limited data, system only takes into account limited features for predicting the resale value of the car. Since this is an online system, current system does not take into account any physical damage to the car body or engine while predicting the resale value The new system developed by us consists of two parts - Data gathering and Prediction using Machine Learning based algorithms.

1.2 Purpose:

The main idea of making a car resale value prediction system is to get hands-on practice for python using Data Science. Car resale value prediction is the system to predict the amount of resale value based on the parameters provided by the user. User enters the details of the car into the form given and accordingly the car resale value is predicted

2. LITERATURE SURVEY:

2.1 Existing problem:

we often feel we need more than 24 hrs. a day to cope up with everything we have in our schedule. Well, that's not possible but reducing the time by changing the conventional method of car resale value prediction can help.

2.2 References:

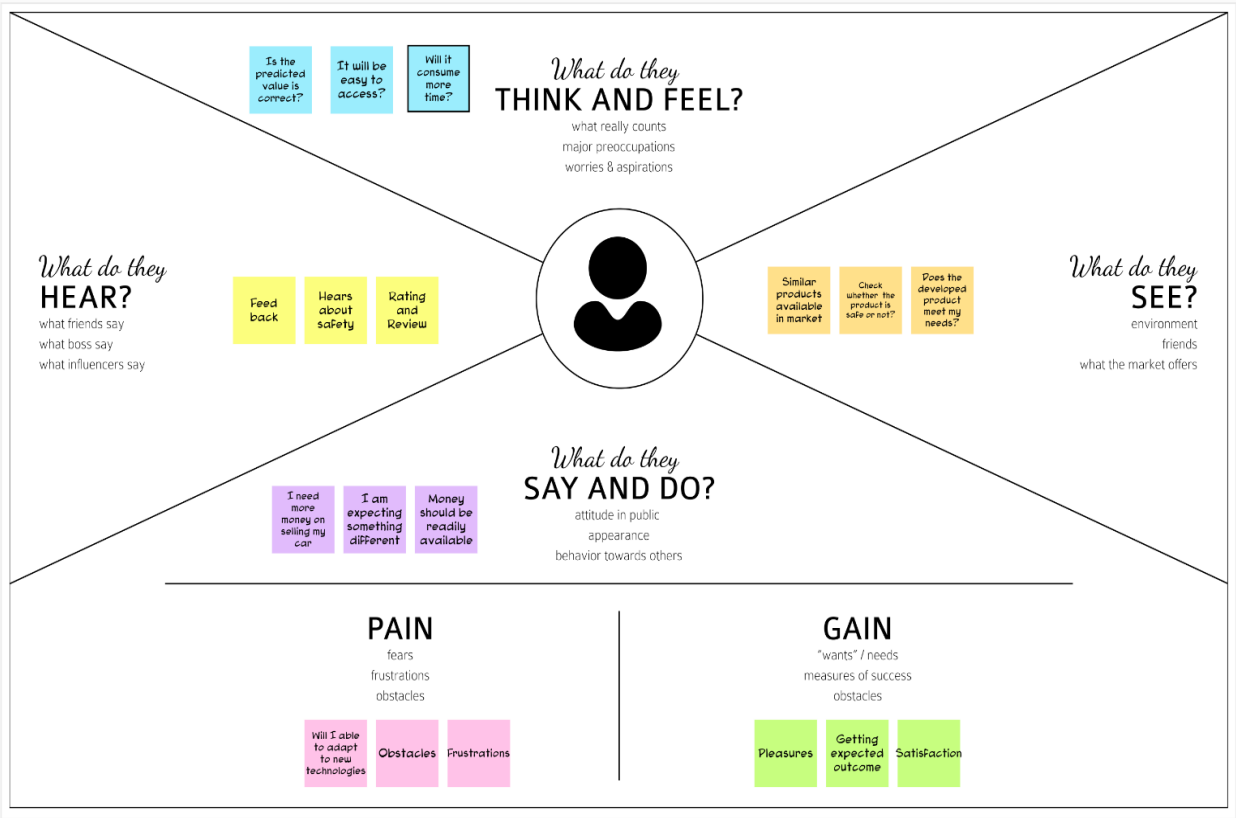
https://www.mural.co/templates/empathy-map-canvas
https://miro.com/templates/customer-problem-statement/
https://www.mural.co/templates/empathy-map-canvas
https://www.ideahackers.network/problem-solution-fit-canvas/
https://medium.com/@epicantus/problem-solution-fit-canvas-aa3dd59cb4fa

2.3 Problem Statement Definition:

The research objective of this study is to predict used cars prices in using data mining techniques, by scraping data from websites that sell used cars, and analysing the different aspects and factors that lead to the actual used car price valuation. To enable consumers to know the actual worth of their car or desired car, by simply providing the program with a set of attributes from the desired car to predict the car price

3. IDERATURE & PROPOSED SOLUTION:

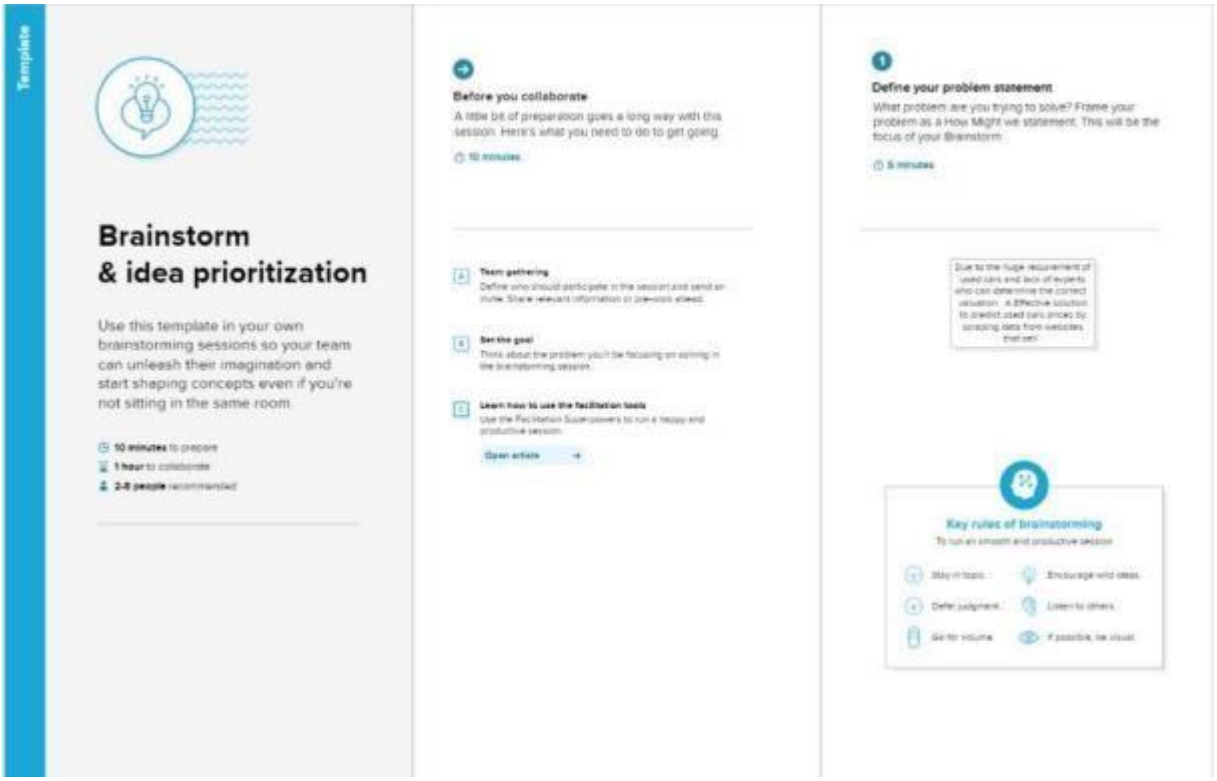
3.1 Empathy Map Canvas:



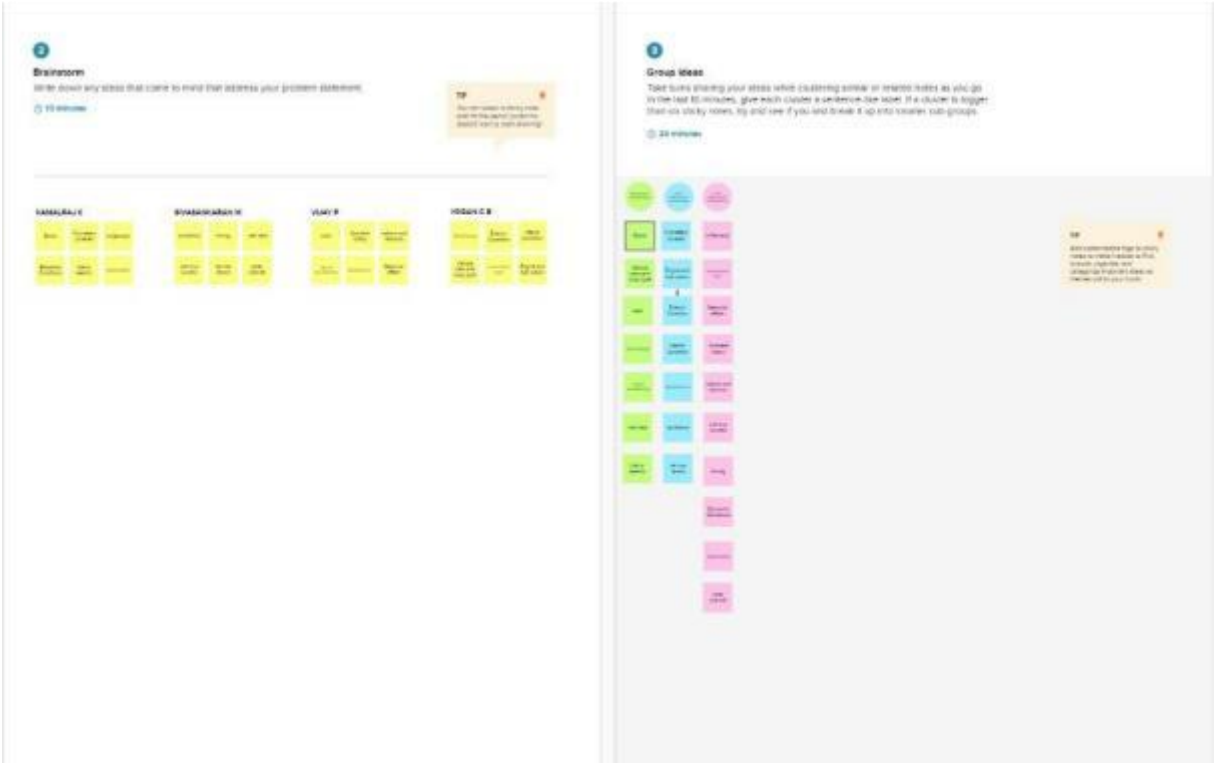
3.2 Brainstorm & Idea Prioritization:

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions. Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

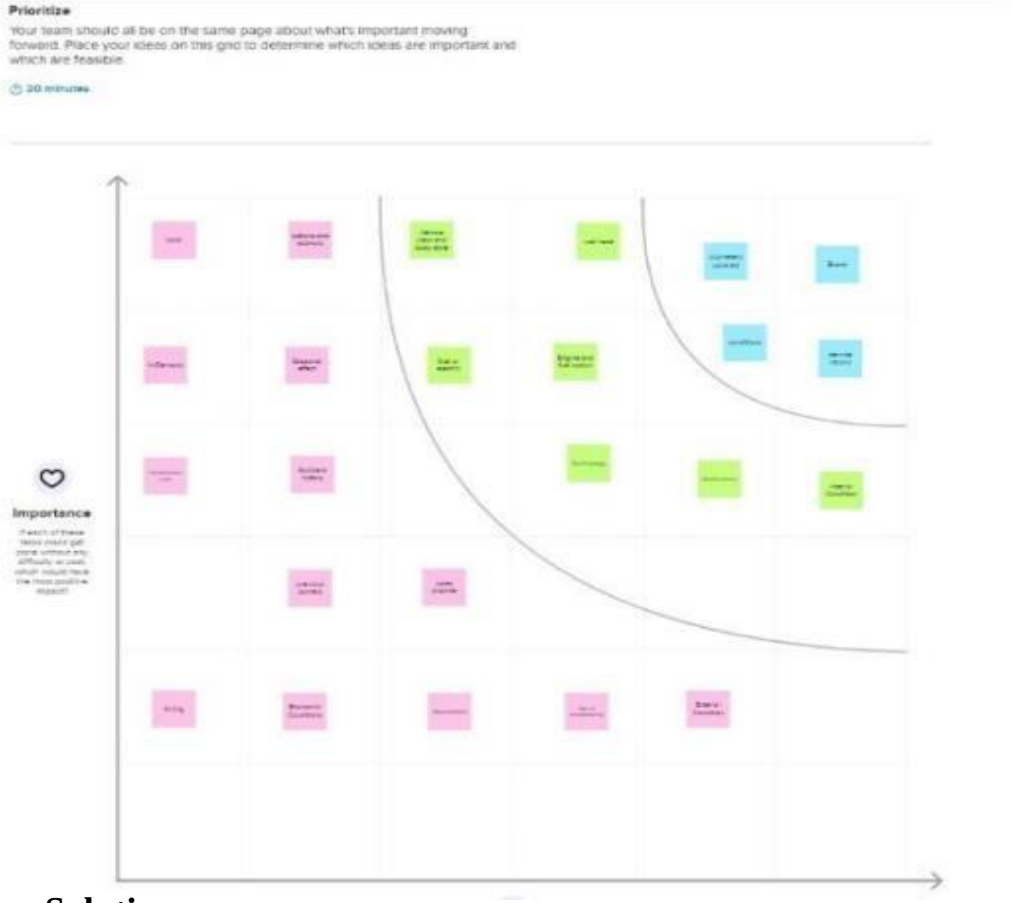
Step-1: Team Gathering, Collaboration and Select the Problem Statemen



tep-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization



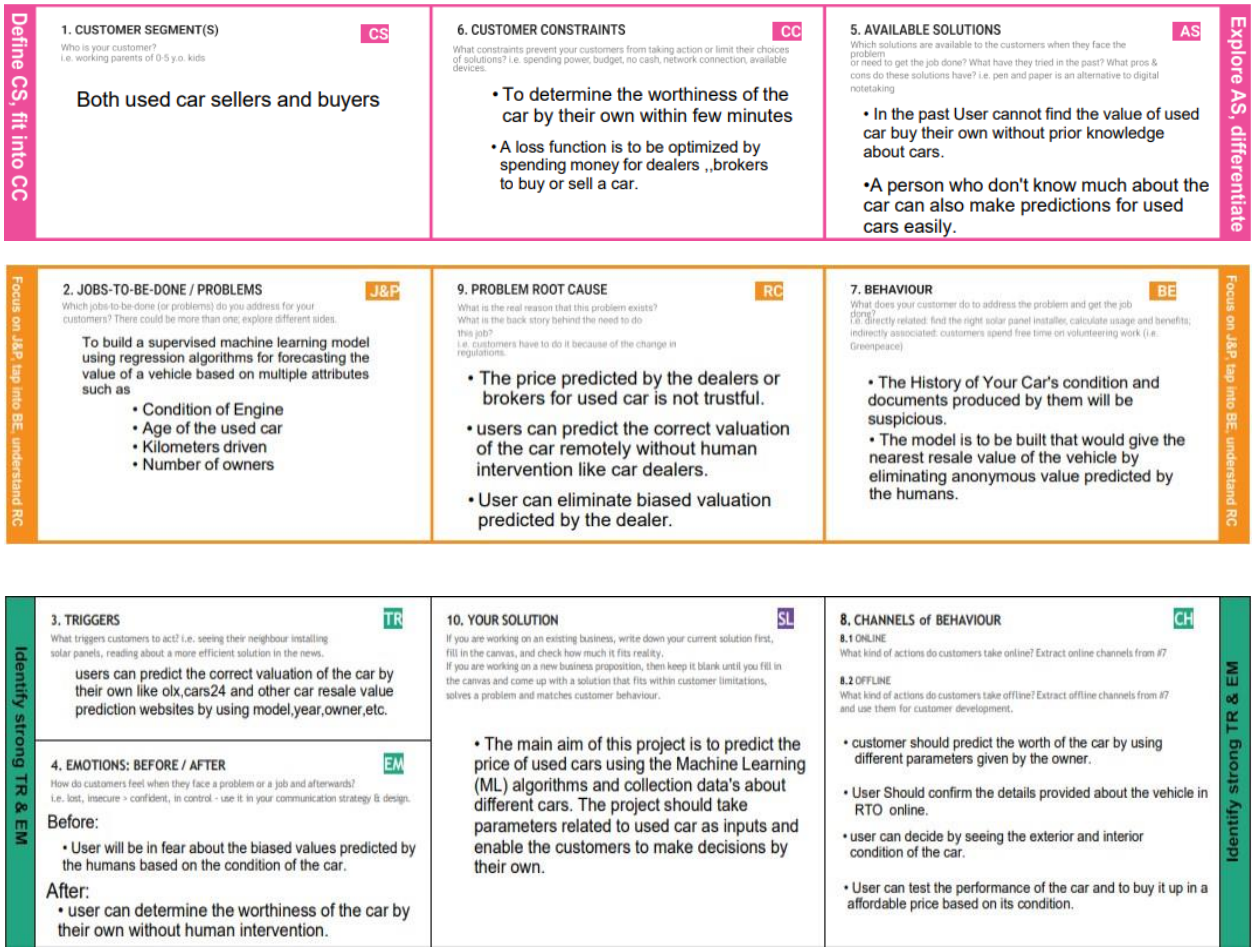
3.3 Propose Solution:

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Sales prediction is the current numerous trend in which all the business companies thrive and it also aids the organization or concern in determining the future goals for it and its plan and procedure to achieve it. ➢ Resale of cars almost occupy a major part in every sales economy. ➢ In that regard various factors like registration year, engine condition, company service record, spare parts condition, tire condition, car body condition, kilometers covered, Interior look, color, mileage, number of owners, battery condition are taken into consideration before buying it along with engine condition and insurance. ➢ The predication using the factors would suggest the final product to be brought. ➢ But these data may be inaccurate at times and there is a need of a proper algorithm that will provide a result with good accuracy rate.

2.	Idea / Solution description	➢The overall proposed idea is to predict the car resale value and showitto the required people. ➢ This idea can beimplemented and could be presented tothecustomer. This involves two phases. ➢ One phase is collecting the dataset for training the car resale value prediction model. ➢ Testing the car resale value prediction model. ➢ The second phase involves creating a website (front end) for presenting the entire solution as a customized GUI so that this would be very useful for the user to utilize this solution. ➢ The user will be asked to enter the details for prediction like model, price, design, kilometers covered, Interior look, color. ➢ If user clicks the predict option, the predicted resale value will be displayed in the
3.	Novelty / Uniqueness	➢Consumer behavior changes, it's a fact. So for better accuracy select a more recently added product when possible. ➢ You can use multiple reference products to get the best average and the novelty sales estimates will be based on features from all of them using the average.
4.	Business Model (Revenue Model)	➢ It helps users to predict the correct valuation of the car remotely perfect valuation and without human intervention like car dealers in the process to eliminate biased valuation predicted by the dealer
5.	Social Impact / Customer Satisfaction	➢ Sales forecasting helps you attain this revenue efficiency by offering insight into the likely behavior of your most valuable customers. ➢ You can predict future sales, as well as improve pricing, advertising, and product development.
6.	Scalability of the Solution	Here we are using time series analysis so, When historical data for a product or product line is available and patterns are obvious, organizations typically employ the time series analysis technique to demand forecasting. A time series analysis can help you detect seasonal variations in demand, cyclical patterns, and major sales trends. The time series analysis approach works best for well-established organizations with several years of data to

		work with and very steady trend patterns.

3.4Problem solution:



4. REQUIREMENT ANALYSIS

4.1 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 Website
FR-2	User Confirmation	Confirmation via website
FR-3	Car Registration	Registering the car details
FR-4	Value Prediction	Predicting the car resale value

4.2 Non-functional Requirements:

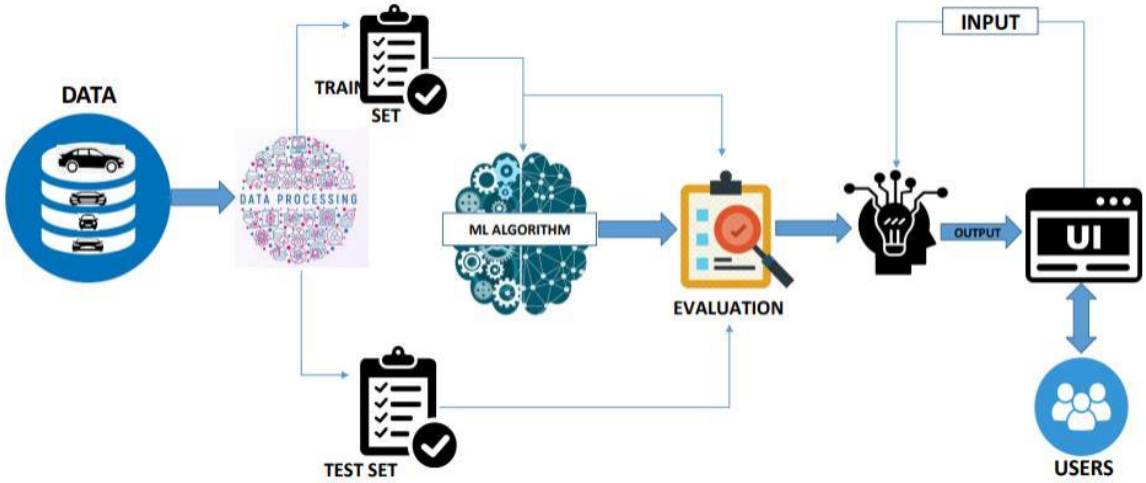
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Predicting the resale value
NFR-2	Security	Providing security to the website
NFR-3	Reliability	Providing high reliability by predicting values for different types of cars
NFR-4	Performance	Providing high performance by using some machine learning techniques
NFR-5	Availability	It is used for all types of cars
NFR-6	Scalability	Predicting values for different types of cars

5. PROJECT DESING:

5.1 Data Flow Diagram

5.2 Solution & Technical Architecture



5.3 User Stories

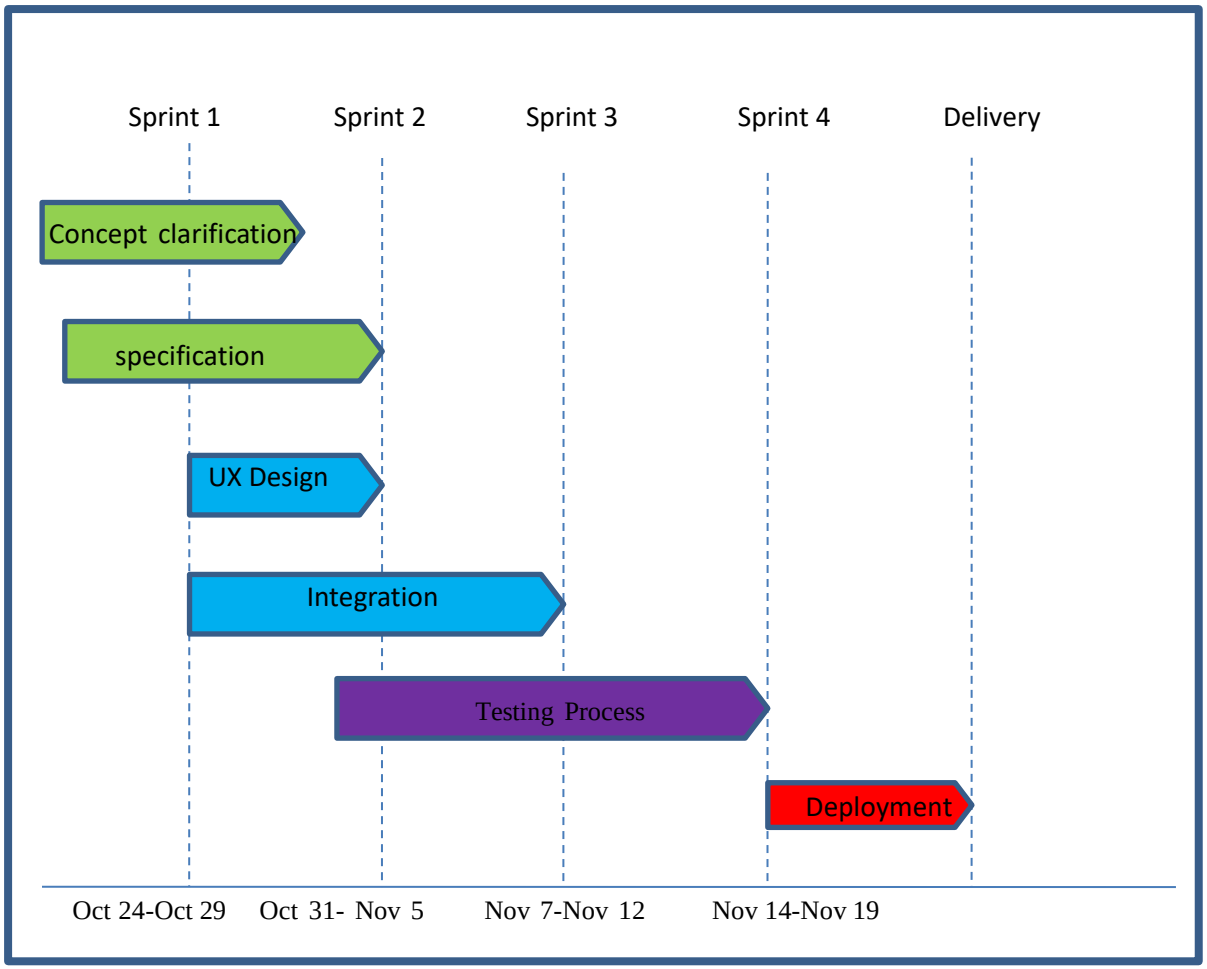
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the car details application by entering my email, password, and confirming my password.	I can access my dashboard and view the car details	High	Sprint-1
		USN-2	As a user, I will receive car resale value in the application.	I can receive my car resale value in the application	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register & access the dashboard with Gmail Login	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	I can access my dashboard and view the car details	High	Sprint-1
	Dashboard	USN-1-5	Show the details of different varieties of used cars.	I can know the resale value of a car	High	Sprint-1

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Milestone Name	Milestone Number	Description		Optional
Pre-Requisites	M-001	We will be downloading the following anaconda software to complete this project and also will be learning some concepts.	Yes	
Prior Knowledge	M-002	We will be learning the supervised learning, unsupervised learning, flask, matrices	Yes	
Project objectives	M-003	We will get the knowledge about the machine learning algorithms, python with machine learning, clean the data, real time analysis of project, building user interface	Yes	
Project flow	M-004	In this installing required libraries, data collection, data preprocessing, model building, application building, final UI	Yes	
Project structure	M-005	We will be building a flask application that needs HTML pages and this model is built in notebook floods	Yes	
Data collection	M-006	Downloading the dataset for the project from the open sources like keggel.com, data.gov	Yes	
Visualizing and analyzing the data	M-007	Importing the important libraries for the project, reading the dataset, univariate, bivariate, multivariate, descriptive analyzing of project done in this phase	Yes	
Data preprocessing	M-008	Finding the shape of the dataset and converting the categorical data to integer encoding or binary encoding and balancing dataset, scaling dataset.	Yes	
Model building	M-009	Model building with the use of four algorithms best algorithm used in the future Decision tree, random forest, KNN, xgboost model are used.	Yes	
Application building	M-010	Building the html pages, python code with all tests done running the application	Yes	
Train the model on IBM	M-011	We will learning to built deep learning and deploying it on the cloud	Yes	
Ideation phase	M-012	Literature survey on the project and preparing the empathy map	Yes	
Project design phase	M-013	Prepare proposed solution, problem solution fit and solution architecture	Yes	
Project design phase 2 Project planning	M-014 M-015	Prepare the customer journey map, functional requirement document, data flow diagrams, technology architecture for the project	Yes Yes	
Project development phase	M-016	Project development delivery of sprint 1, sprint 2, sprint 3, sprint 4	Yes	

6.2 Sprint Delivery Scheduling



6.3 Reports from JIRA

---

7. CODING & SOLUTION

```
from flask import Flask,render_template,request
import jsonify
import requests
import pickle
import numpy as np
import sklearn

from sklearn.preprocessing import StandardScaler
```

```
# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "nsMPxkqMjnLbFOa4okSPCk2ZKV0Mr8Mry0NxF6J3KgUb"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]
header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}


app = Flask(__name__)
model = pickle.load(open('file.pkl','rb'))

@app.route('/',methods=['GET'])
def Home():
  return render_template('index.html')


standard_to = StandardScaler()

@app.route('/predict',methods = ['POST'])
def predict():
    Fuel_Type_Diesel =0
    if request.method == 'POST':
        Year = int(request.form['Year'])
        Present_Price = float(request.form['Present_Price'])
```

```
if request.method == 'POST':
    Year = int(request.form['Year'])
    Present_Price = float(request.form['Present_Price'])
    Kms_Driven = int(request.form['Kms_Driven'])
    Owner = int(request.form['Owner'])
    Fuel_Type_Petrol = request.form['Fuel_Type_Petrol']
    if(Fuel_Type_Petrol == 'Petrol'):
        Fuel_Type_Diesel = 0
        Fuel_Type_Petrol = 1

    elif(Fuel_Type_Diesel=='Diesel'):
        Fuel_Type_Petrol = 0
        Fuel_Type_Diesel = 1
    else:
        Fuel_Type_Petrol = 0
        Fuel_Type_Diesel = 0

    Year = 2020 - Year
    Seller_Type_Individual = request.form['Seller_Type_Individual']
    if(Seller_Type_Individual=='Individual'):
        Seller_Type_Individual =1
    else:
        Seller_Type_Individual = 0

    Transmission_Manual = request.form['Transmission_Manual']
    if(Transmission_Manual == 'Manual'):
        Transmission_Manual = 1
    else:
        Transmission_Manual = 0

    prediction = model.predict([[Present_Price,Kms_Driven,Owner,Year,Fuel_Type_Diesel,Fuel_Type_Petrol,Selle
    output = round(prediction[0],2)
```

```
        prediction = model.predict([[Present_Price,Kms_Driven,Owner,Year,Fuel_Type_Diesel,Fuel_Type_Petrol,Selle
        output = round(prediction[0],2)


        feilds=[Present_Price,Kms_Driven,Owner,Year,Fuel_Type_Diesel,Fuel_Type_Petrol,Seller_Type_Individual,Tra
        payload_scoring = {"input_data": [{"fields": [['Present_Price','Kms_Driven','Owner','Year','Fuel_Type_Di
        response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/1a9cf846-2d3f-459a
        pred=response_scoring.json()
        out=pred['predictions'][0]['values'][0][0]


        if output<0:
            return render_template('index.html',prediction_text='Sorry! You cannot sell this car')
        else:
            return render_template('index.html', prediction_text='You can sell this car at Rs.{} lakhs'.format(o

    else:
        return render_template('index.html')



if __name__ == '__main__':
    app.run(debug=True)
```

8. TEASTING

8.1 Test Cases

• Missing values

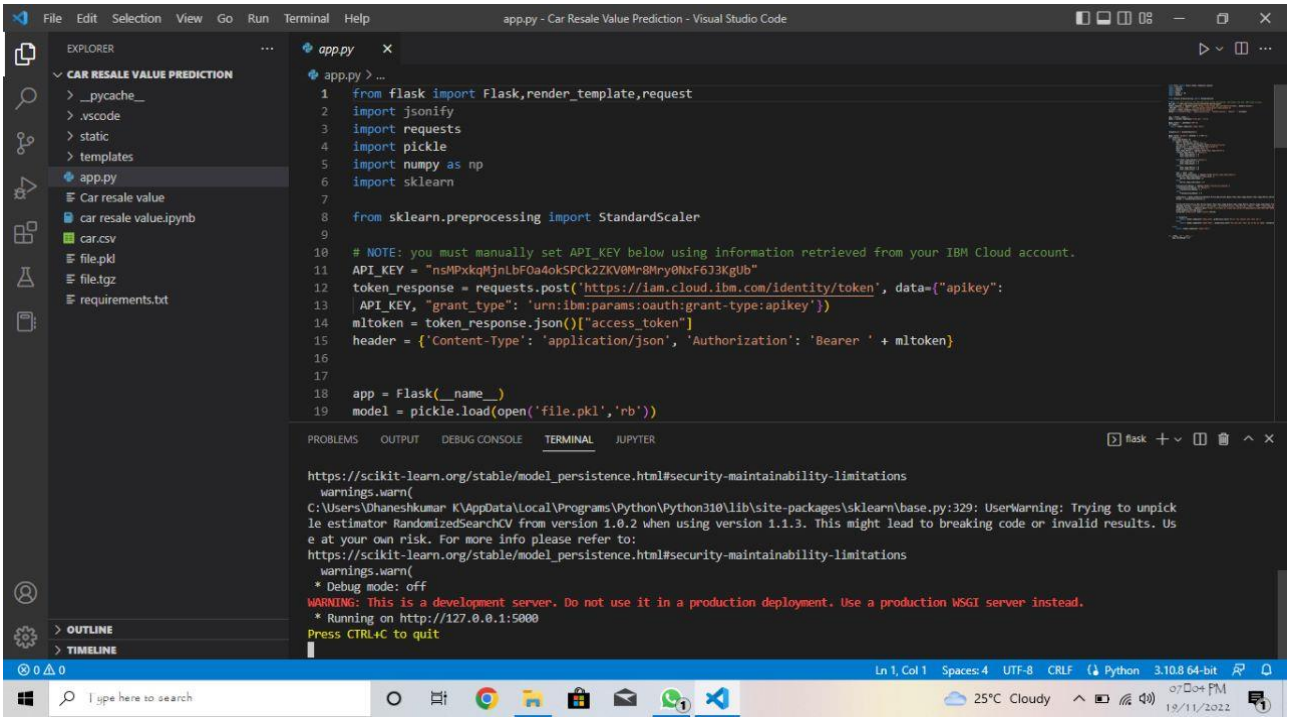
The trained ML model requires 4 feature inputs for predicting the output. Failing which, the model throws invalid Input error. All the fields in the html form have been marked required using CSS and thus user must input all fields.

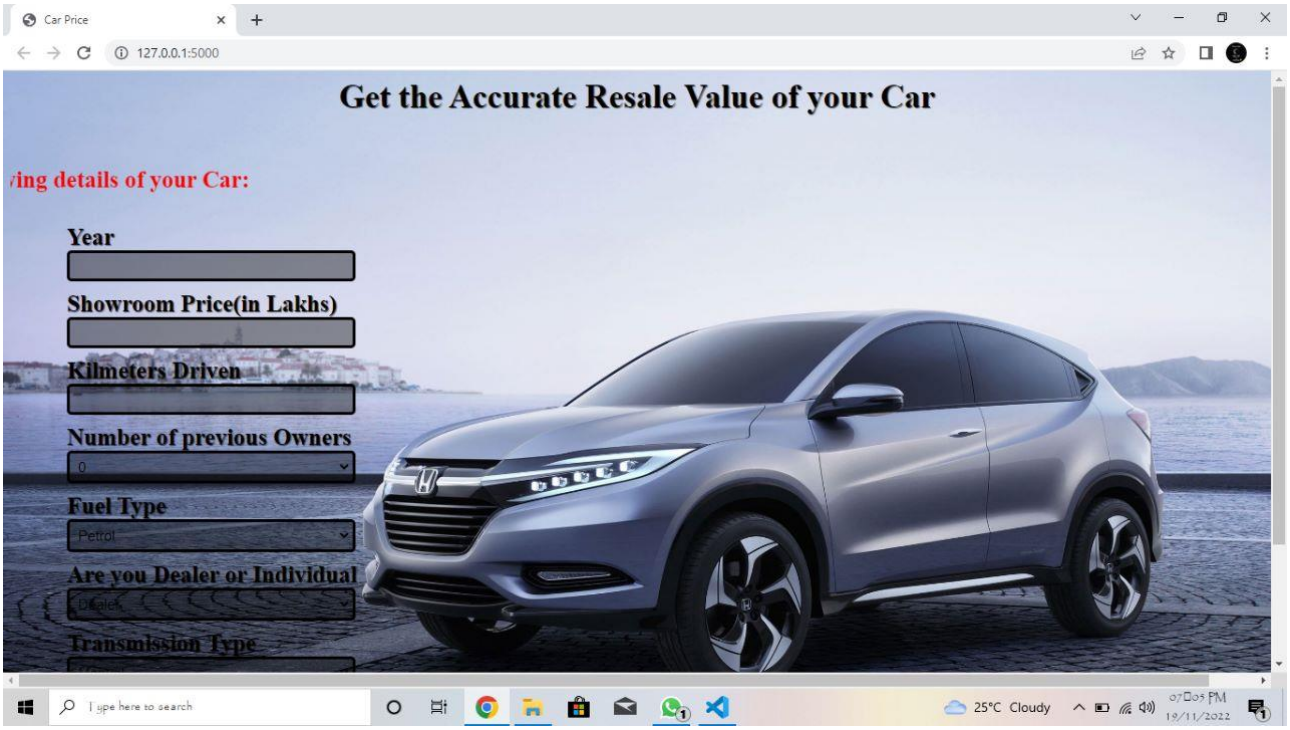
Output: User must input all the fields, failing which, form shows warning message "this field needs to be filled". Thus, there can be no errors in model prediction.

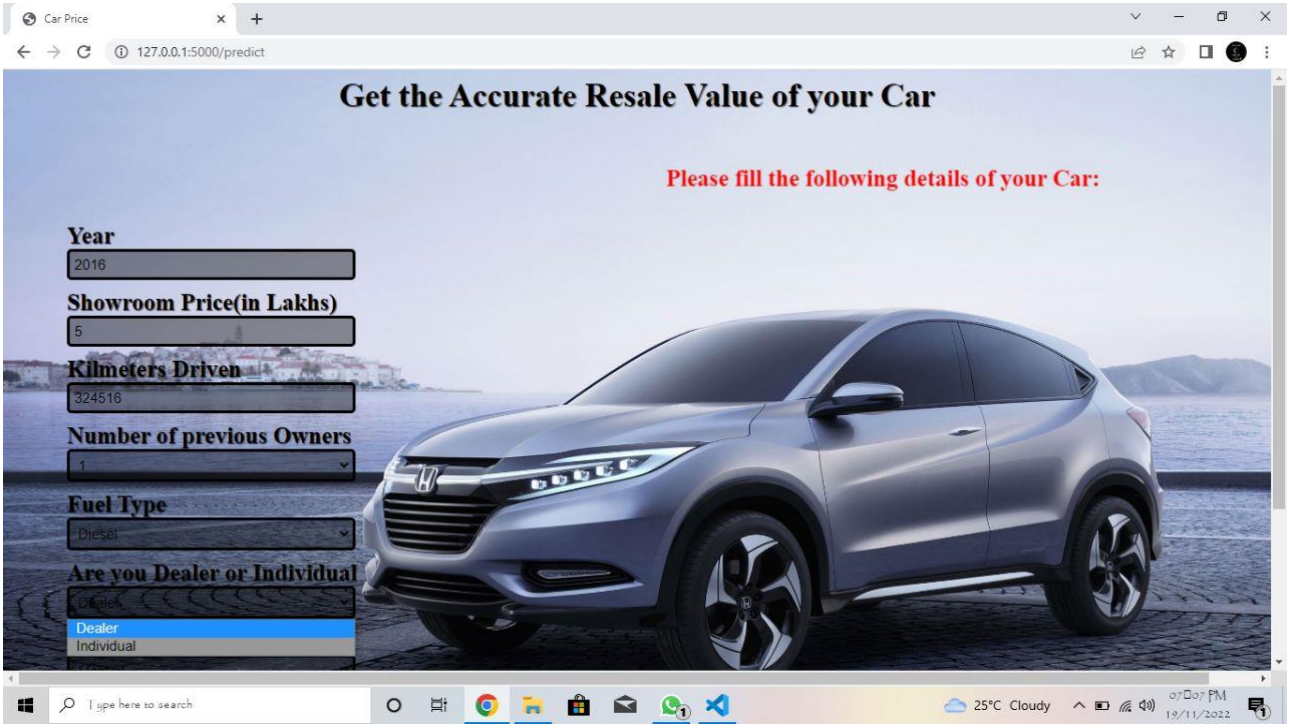
• Invalid Input

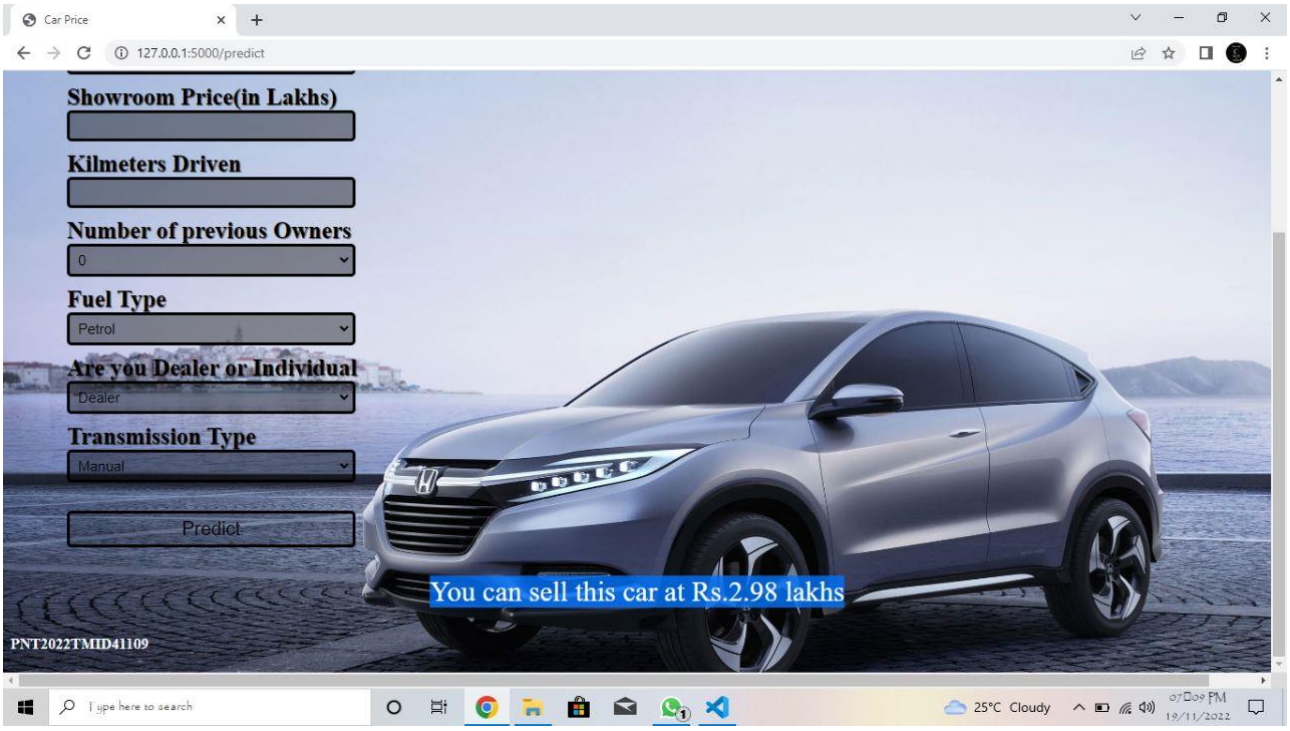
The trained ML model requires only numerical input for all 4 features. Thus, if user uses symbols such as comma while input, model may throw error. To overcome the same, preprocessing script is deployed in backend which removes all unwanted characters like comma, whitespaces etc. so that model gets required input

9. RESULTS


app.py - Car Resale Value Prediction - Visual Studio Code
CAR RESALE VALUE PREDICTION
app.py
Car resale value
car resale value.ipynb
car.csv
file.pkl
file.tgz
requirements.txt
from flask import Flask,render_template,request
import jsonify
import requests
import pickle
import numpy as np
import sklearn
from sklearn.preprocessing import StandardScaler
# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "nsMPxkqMjnLbFOa4okSPCk2ZKV0Mr8Mry0NxF6J3KgUb"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]
header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
app = Flask(__name__)
model = pickle.load(open('file.pkl','rb'))
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
JUPYTER
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
C:\Users\Dhaneshkumar K\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpick
le estimator RandomizedSearchCV from version 1.0.2 when using version 1.1.3. This might lead to breaking code or invalid results. Us
e at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
* Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
OUTLINE
TIMELINE
Ln 1, Col 1
Spaces: 4
UTF-8
CRLF
Python
3.10.8 64-bit
25°C Cloudy


Car Price
127.0.0.1:5000
Get the Accurate Resale Value of your Car
ving details of your Car:
Year
Showroom Price(in Lakhs)
Kilometers Driven
Number of previous Owners
0
Fuel Type
Petrol
Are you Dealer or Individual
Dealer
Transmission Type
25°C Cloudy


Car Price
127.0.0.1:5000/predict
Get the Accurate Resale Value of your Car
Please fill the following details of your Car:
Year
2016
Showroom Price(in Lakhs)
5
Kilometers Driven
324516
Number of previous Owners
1
Fuel Type
Diesel
Are you Dealer or Individual
Dealer
Individual
Type here to search
25°C Cloudy
07:07 PM
19/11/2022


Car Price
127.0.0.1:5000/predict
Showroom Price(in Lakhs)
Kilometers Driven
Number of previous Owners
0
Fuel Type
Petrol
Are you Dealer or Individual
Dealer
Transmission Type
Manual
Predict
You can sell this car at Rs.2.98 lakhs
PNT2022TMID41109
Type here to search
25°C Cloudy
07:09 PM
19/11/2022

10. ADVANTAGES & DISADVANTAGES

Advantages:

- Good at learning complex and non-linear relationships
- Highly explainable and easy to interpret
- Robust to outliers
- No feature scaling is required

Disadvantages:

- Consumes more time
- Require high computational power

11. CONCLUUSION:

The increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on a global increase. Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of used car price prediction. This paper compares 3 different algorithms for machine learning : Linear Regression, Lasso Regression and Ridge Regression

12. FUTURE SCOPE:

In future this machine learning model may bind with various website which can provide real time data for price prediction. Also we may add large historical data of car price which can help to improve accuracy of the machine learning model. We can build an android app as user interface for interacting with user. For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole datase

13.APPENDIX:

GitHub

https://github.com/IBM-EPBL/IBM-Project-41824-1660645243

Project demo link

Our project runs on local host we can't share or use the site
we attached source code through the link below

https://github.com/IBM-EPBL/IBM-Project-41824-1660645243/tree/main/Project%20Final%20Deliverable/Car%20Resale%20Value%20Prediction

https://github.com/IBM-EPBL/IBM-Project-41824-1660645243/tree/main/Project%20Final%20Deliverable/Demo_Vedio