PROJECT REPORT

PROJECT TITLE: CAR RESALE PRICE PREDICTION

TEAM ID: PNT2022TMID1436

TEAM MEMBERS:

NANDHINI S
PALADUGU SNEHITHA
PREETHI D
RAGAVI VS

1.INTRODUCTION

1.1Project Overview

With difficult economic conditions, it is likely that sales of second-hand 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.

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

2.1Existing problem

With some exceptions, cars tend to lose their value rapidly. They can drop by 20-30% in the first year and 60% or more of the original price after five years. The curve flattens out from there, but until and unless your car becomes a collectible or a classic, the value continues to go down. In many cases, the demand for your vehicle on the used car market also decreases. Newer trade-ins typically bring higher prices and bigger profits for the dealer.Low mileage can tip the balance in favor of an older trade-in. According to the U.S. Department of Transportation, the average miles driven in the United States is roughly 13,500 a year. Anything under the average is a selling point for the dealer looking to make a profit reselling your trade. A 10 year-old car with 67,500 miles on it (half the national average) could be more attractive to potential buyers than one with 135,000 miles or more.Regardless of age or mileage, condition plays a huge part in used car trade-in value. Dealers may not look at scratches, dents, stained interiors, cracked windshields, and mechanical issues and think "stuff happens." They look at what it will cost them to repair or replace the things necessary for the car to sell at a strong price. Additionally, they'll deduct that from what otherwise could have been a decent offer for your trade-in.Beyond equipment, there are other factors that can vary. An attractive, somewhat neutral paint color can appeal to a larger number of used car customers. A model that sold poorly or a make that's gone out of business could be at a disadvantage. Likewise, demand for convertibles is likely to be much stronger in the Sunbelt than the Snowbelt. Also, cars with good gas mileage can be worth more in regions with higher fuel prices.Before you head down to the dealership, clean the car. At minimum, vacuum it out, remove your personal items and run it through a car wash. A full detailing of the exterior, interior and engine bay might be an even better idea. It conveys a well-cared for car and could easily pay for itself in trade-in value.Ultimately, getting the best used car trade-in value is a process that should begin when you buy that car. Look for one with historically strong resale value in a popular color with popular equipment. Keep the mileage below average

if you can. Stick to the manufacturer's recommended maintenance schedules, keep the receipts, and fix anything that breaks right away. Finally, monitor your car's value. Wait for the right moment when you have enough equity in your old car to make a difference in the negotiation for your new one.

2.2 References

[1] Pudaruth, S., 2014. “Predicting the Price of Used Cars using Machine Learning Techniques.” Vol 4, Number 7 (2014), pp. 753-76. 2).

[2] Gokce, E. (2020, January 10). “Predicting used car prices with machine learning techniques. “

[3] Predicting Used Car Prices with Machine Learning Techniques | by Enes Gokce | Towards Data Science

[4] Ning sun, Hongxi Bai, Yuxia Geng, Huizhu Shi, “Price Evaluation Model In Second Hand Car System Based On BP Neural Network Theory”; (Hohai University Changzhou, China)

[5] Nitis Monburinon, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, Pitchayakit Boonpou, “Prediction of Prices for Used Car by using Regression Models” (ICBIR 2018)

2.3 Problem Statement Definition

This resale value prediction system is made for general purpose to just predict the amount that can be roughly acquired by the user. We try to predict the amount of resale by best 70% accuracy so the user can get estimated value before he resales the car and doesn't make a deal in loss.

3.IDEATION & PROPOSED SOLUTION

3.1Empathy Map Canvas

3.2 Ideation & Brainstorming

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

Brainstorm & idea prioritization

Car Resale Value Prediction

10 minutes to prepare
1 hour to collaborate
4 people

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

A **Team gathering**
Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

B **Set the goal**
Think about the problem you'll be focusing on solving in the brainstorming session.

C **Learn how to use the facilitation tools**
Use the Facilitation Superpowers to run a happy and productive session.

Open article

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM
How might we [your problem statement]?

Key rules of brainstorming

To run an smooth and productive session

Stay in topic.
Encourage wild ideas.
Defer judgment.
Listen to others.

Go for volume. If possible, be visual.

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!

Nandhini

- Engine Type
- Insurance
- Transciever
- Suspension

Preethi

- Gear type
- Steering
- Engine mode
- Ownership

Ragavi

- Front axle
- AI based
- Brake System
- Blue type

Snehitha

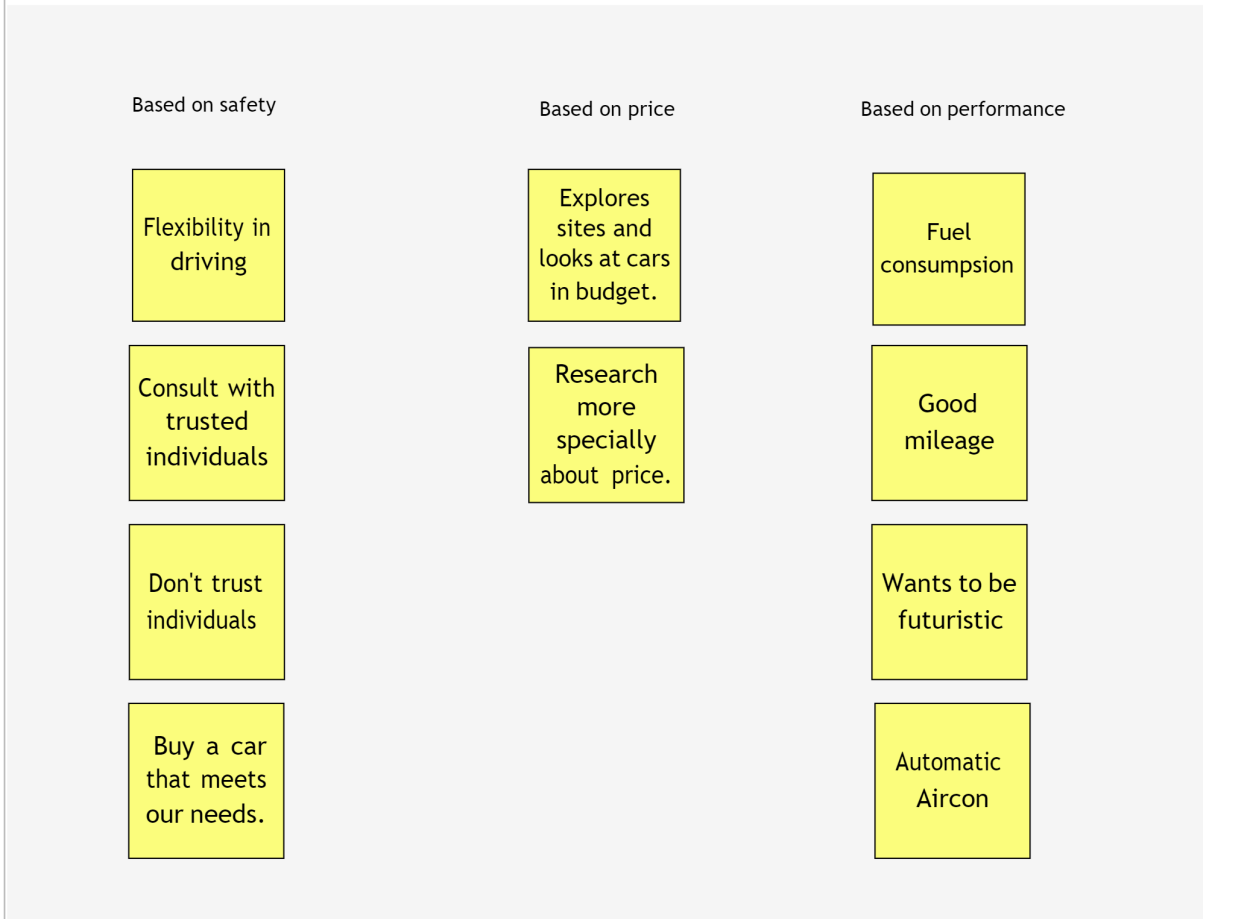
- Fitness certificate
- Fuel resource
- Mother board
- Alternator

3

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



4

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

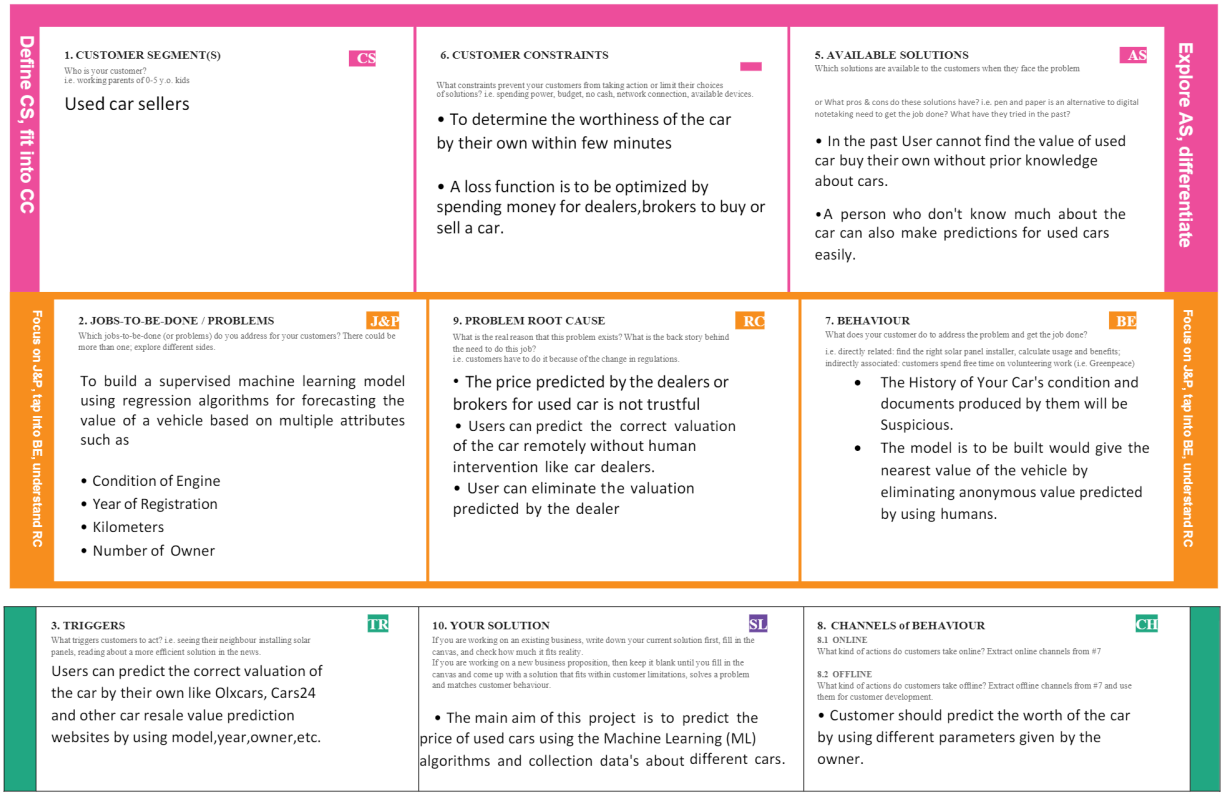
3.3 Proposed Solution

Proposed Solution :

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To develop a webpage to predict the resale value of a car
2.	Idea / Solution description	To train the system with the dataset using a regression model
3.	Novelty / Uniqueness	By using the optimal regression model to predict the value in a less amount to time and predict its value
4.	Social Impact / Customer Satisfaction	The customer can get an idea about the resale value of their car .to have an idea whether to sell their vehicle or not based on their financial condition
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 solution given by the trained system is efficient and is nearly accurate value of the vehicle.

3.4 Problem Solution Fit

Project Title: Car Resale Value Prediction | **Project Design Phase-I - Solution Fit Template** | **Team ID:** IBMSI20220014364



4.REQUIREMENT ANALYSIS

4.1 Functional requirement

FUNCTIONAL REQUIREMENTS
ID: IBMSI20220014364

Anaconda Navigator :

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and macOS. Conda is an open-source, cross-platform, package management system. Anaconda comes with great tools like JupyterLab, Jupyter Notebook, QtConsole, Spyder, Glueviz, Orange, Rstudio, Visual Studio Code.

For this project, we will be using **Jupyter notebook and Spyder**

To install Anaconda navigator and to know how to use Jupyter Notebook & Spyder using Anaconda watch the video

To build Machine learning models you must require the following packages

Sklearn: Scikit-learn is a library in Python that provides many unsupervised and supervised learning algorithms.

NumPy: NumPy is a Python package that stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object

Pandas: pandas is a fast, powerful, flexible, and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.

Matplotlib: It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits

Flask: Web framework used for building Web applications.

1. Open anaconda prompt.
2. Type "pip install numpy" and click enter.
3. Type "pip install pandas" and click enter.
4. Type "pip install matplotlib" and click enter.
5. Type "pip install scikit-learn" and click enter.
6. Type "pip install Flask" and click enter.

4.2 Non-Functional requirements

Non-functional Requirements:

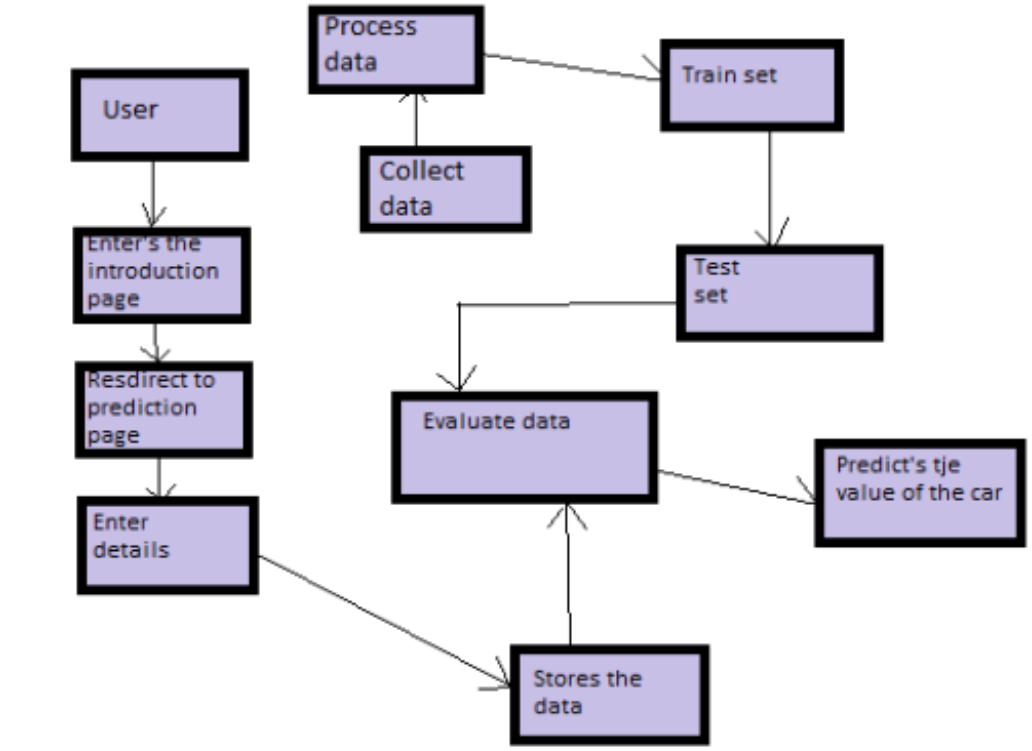
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The predictor must be easy to use and the UI should be smooth and decluttered.
NFR-2	**Security**	It should be ensured that necessary security features are in place to safe guard users' data from activities like data theft
NFR-3	**Reliability**	The reliability of the predictor must be maintained by providing the customer close-to-accurate results every single time.
NFR-4	**Performance**	The performance of the predictor is entirely dependent on its accuracy and the time taken by it to come up with the results.
NFR-5	**Availability**	It must be made accessible through any browsers to ensure that it is available to a wide spectrum of users.
NFR-6	**Scalability**	The predictor must be designed in such a way that its range/scope can easily be increased without any massive changes
NFR-7	**Serviceability**	Customer service must be provided through chat box/chat bots to resolve any issues that they might face and to resolve their queries.
NFR-8	**Manoeuvrability**	The platform must be easily manoeuvrable.

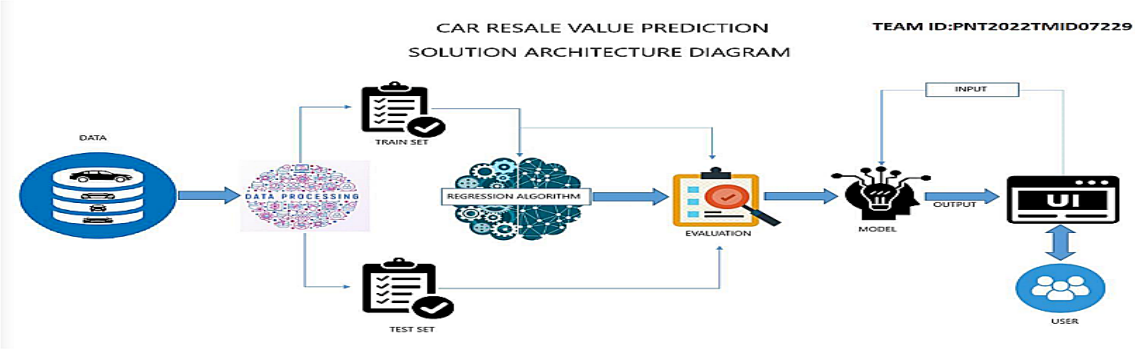
5.PROJECT DESIGN

5.1 Data Flow Diagrams

Data Flow Diagram:



5.2 Solution&TechnicalArchitecture



TECHNICAL ARCHITECTURE

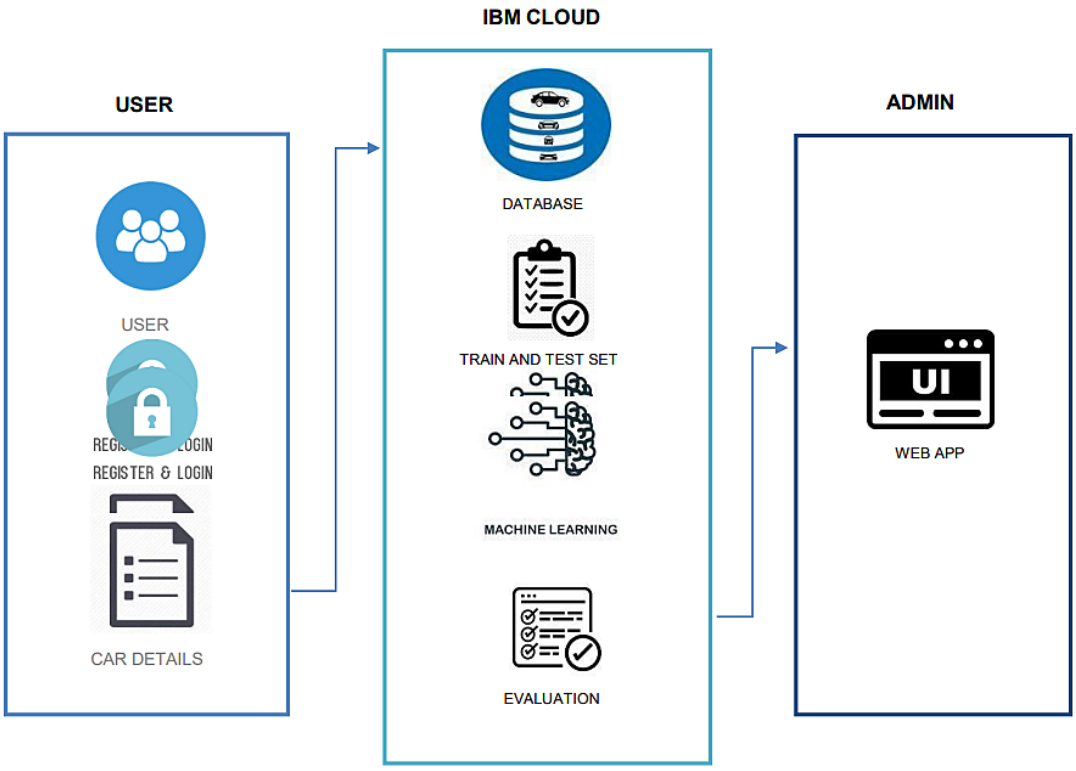


Table-1: Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	A website where the user interacts	HTML, Python ,Flask
2.	Data Pre-process	Pre-processing of the dataset	Pandas, Numpy
3.	Prediction	Resale value Prediction	Python models
4.	Cloud Database	Database Service on Cloud	IBM Cloud
5.	Machine Learning Model	ML model for predicting prices	Regression Model

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	open-source frameworks used for development	Python, Flask, Python Libraries
2.	Security Implementations	User Authentication	SHA-256 Encryption
3.	Scalable Architecture	The 3-tier architecture (Web, Application, Database) is scalable.	IBM Cloud
4.	Availability	Application made available even under heavy load	IBM Cloud - cloud hosting
5.	Performance	Able to handle multiple requests	IBM Cloud - load balancers

5.3 UserStories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (web user)	Enters the browser	USN-1	As a user, I can access to website using a web browser	I can enter by selecting the appropriate web link	High	Sprint-1
		USN-2	As a user, I can proceed to the prediction page by selecting the check value button in the home page	I can enter into it without any acceptance	High	Sprint-1
Customer (mobile user)	Enters into a mobile browser	USN-3	As a user, I can use any of the appropriate mobile browser to enter into the website	I can enter by using an appropriate web link	Medium	Sprint-2
Customer Care Executive	Browser page	USN -4	Customer stuck with any problem, I will assist.	I can help with problems when needed.	Medium	Sprint-3
Administrator	Website	USN -5	Help in the smooth running of the website.	I help running the website.	High	Sprint -4

6. PROJECTPLANNING&SCHEDULING

6.1SprintPlanning&Estimation

Sprint	Functional (Requirements)	User Story Number	User story/Task	Story points	Priority
Sprint 1	Home page	USN 1	As a user, I can view the home page of the web application.	20	Low
Sprint 2	Car resale value display	USN 2	As a user, I can be redirected to the data entry page.	20	medium
Sprint 3	Required data entry	USN 3	As a user, I can enter my car details in the required fields.	20	medium
Sprint 4	Resale vale prediction	USN 4	As a user, I expect the application to predict the resale value of my car.	20	medium

6.2 **SprintDelivery Schedule**

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring to technical papers,research publications etc.	**22th SEPTEMBER 2022**
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements .	**20th SEPTEMBER 2022**
Ideation	List out them by organising the brainstorming session and prioritise the top 3 ideas based on the feasibility & importance.	**18th SEPTEMBER 2022**
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	**25th SEPTEMBER 2022**
Problem Solution Fit	Prepare problem - solution fit document.	**29th SEPTEMBER 2022**
Solution Architecture	Prepare a solution for selected project architecture document.	**01 OCTOBER 2022**

6.3 Reports from JIRA

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.00
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74

Fig 2: Sample data before data preprocessing

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Name	Location	Fuel_Type	Transmission	Owner_Type
0	2010	72000.0	26.60	998.0	58.16	5.0	1.75	18	9	0	1	0
1	2015	41000.0	19.67	1582.0	126.20	5.0	12.50	10	10	1	1	0
2	2011	46000.0	18.20	1199.0	88.70	5.0	4.50	9	2	3	1	0
3	2012	87000.0	20.77	1248.0	88.76	7.0	6.00	18	2	1	1	0
4	2013	40670.0	15.20	1968.0	140.80	5.0	17.74	1	3	1	0	2

Fig 3: Sample data after data preprocessing

7.CODING & SOLUTIONING

7.1 Feature 1

1. IBM Watson Platform
2. Web UI
3. Python Code
4. HTML
5. CSS
6. JS

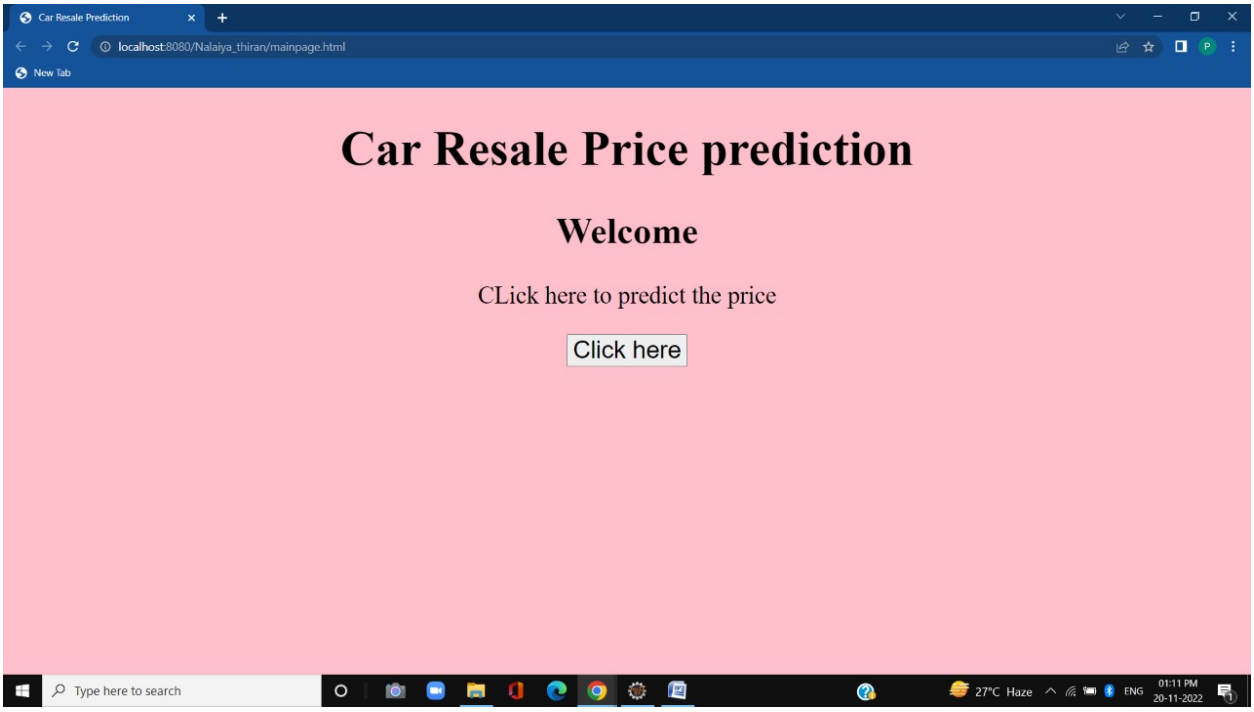
7.2 Feature 2

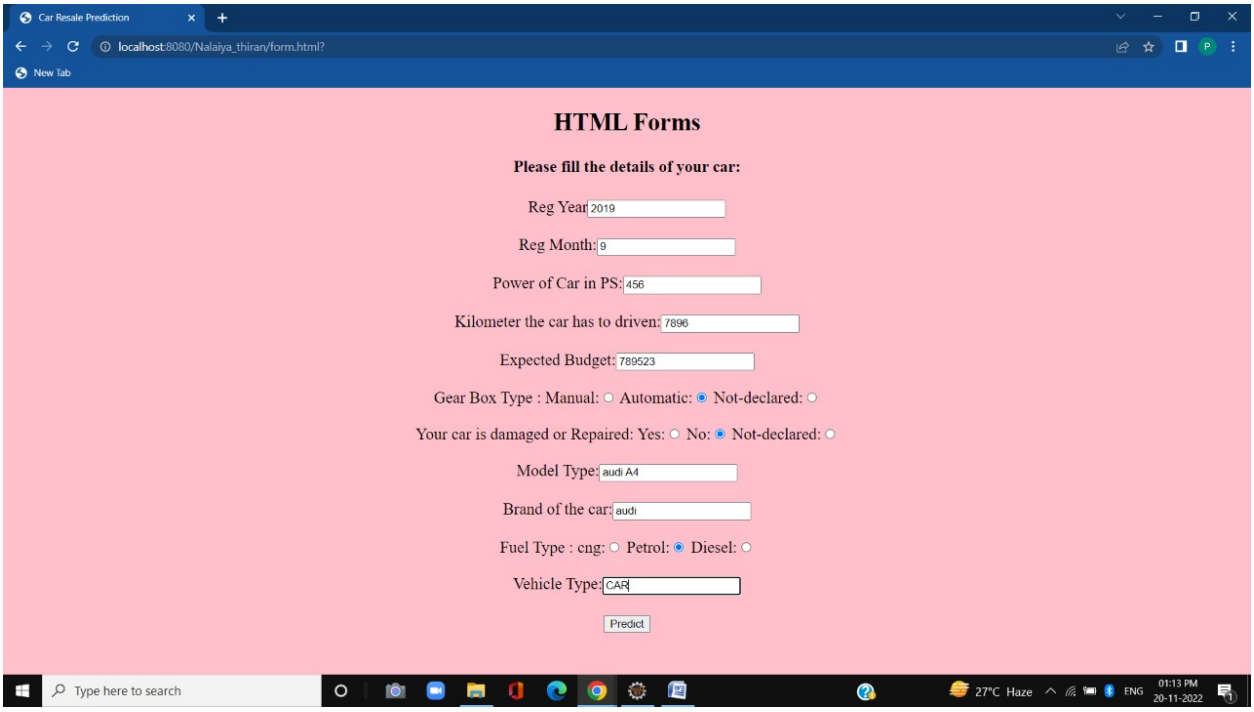
7. Index
8. Chance

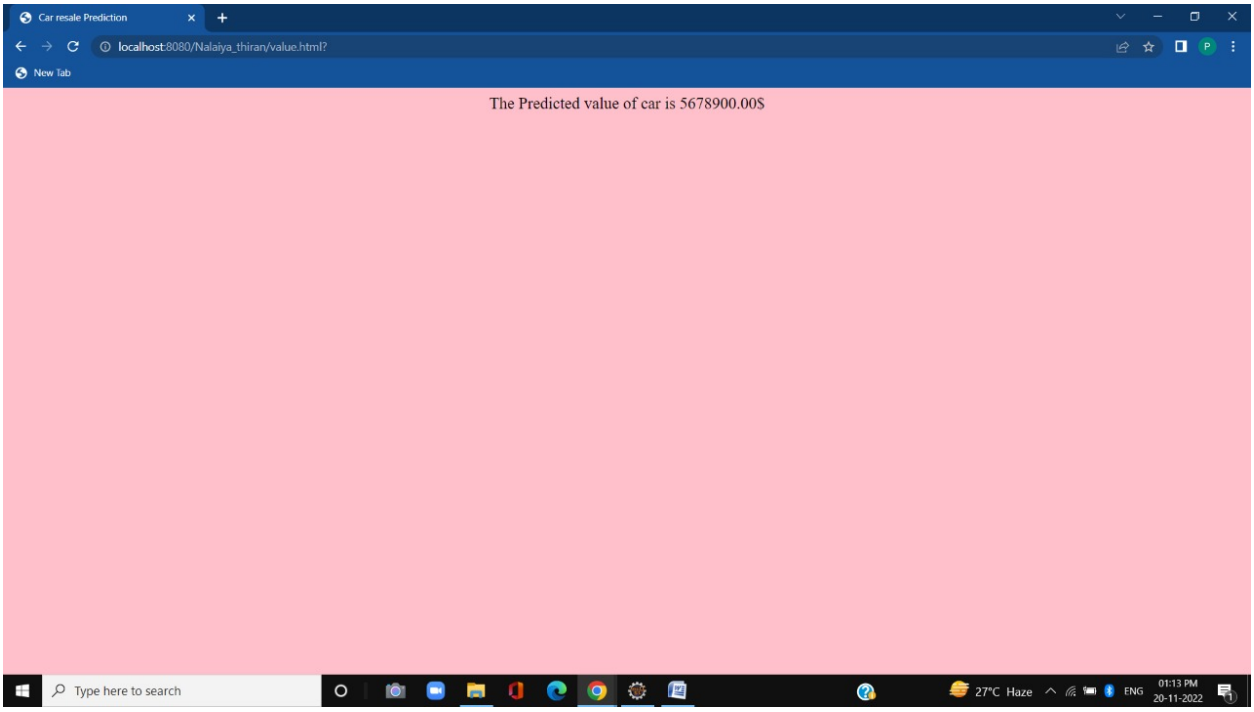
9. Nochance
10. Demo2

8.Testing :

8.1 Test cases







9.RESULTS

9.1Performance metrics

#predicting the values to test

set y_pred = regressor.predict(X_test)

```
#printing the accuracy for test set
        print(r2_score(Y_test,y_pred)
```

10.ADVANTAGES AND DISADVANTAGES

10.1 Advantages

- Highly explainable and easy to interpret
- Robust to outliners
- No feature scaling is required

10.2 Disadvantage

- Consumes more time
- Requires high computational power

11.CONCLUSION

However, once more data is collected and various different cars are included in the system, deep learning-based ANN or LSTM would perform better. But currently, GBR based car valuation system can predict resale value of a car with Root Mean Squared Error (RMSE) of 50,000 INR.

12.FUTURE SCOPE

Currently, system can only deal with Swift Dzire cars due to lack of data. Also, data has been collected of only 5 cities of India. This can be extended to multiple car models and cities so as to improve accuracy and usability.
Efficient use of deep learning such as LSTM (Long shortterm memory) or RNN be implemented once enough data is collected. This can improve accuracy and decrease RMSE drastically.
Currently, only few features are used to predict resale value of the car. This can be extended to more features.
One can also implement CNN to determine physical condition of the car from images like identifying dents, scratches etc. and thus predicting more relevant resale value of a car.

13.APPENDIX

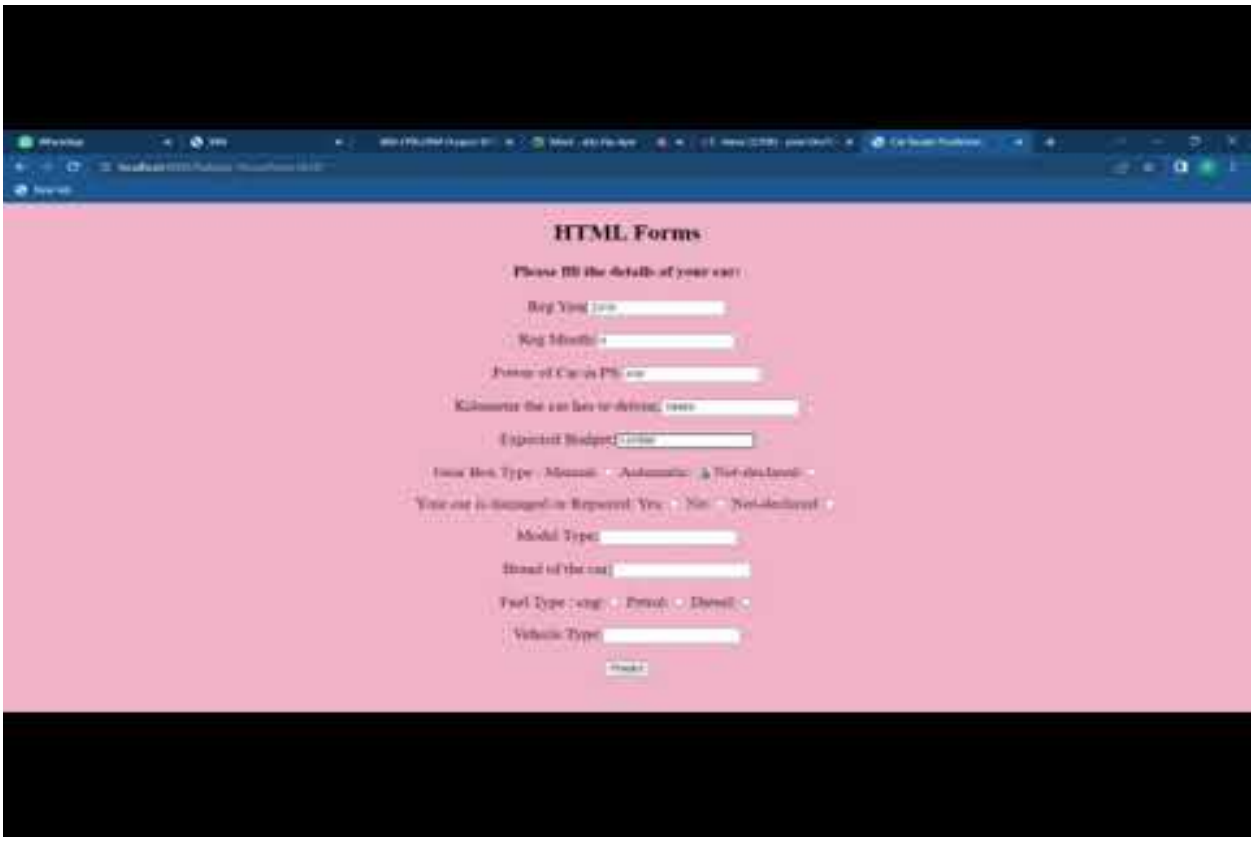
```
#Importing required libraries
 import pandas as pd
Import numpy as np
from flask import Flask, render_template, Response, request
import pickle
from sklearn.preprocessing import LabelEncoder
import pickle



#Load the model and initialize Flask app
app=Flask(__name__)
 filename='resale_model.sav'
model_rand=pickle.load(open(filename,'rb'))


#Configure app.py to fetch the parameter values from the ui,and return the prediction.

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

@app.route(y_predict', methods=['GET', 'POST'])
 def y_predict():

regyear = int (request.form['regyear'])
powerps = float(request.form['powerps'])
 kms = float(request.form['kms'])
 regmonth int(request.form.get('regmonth'))
gearbox = request.form['gearbox']
damage request.form['dam']
```

```
        model request.form.get('modeltype') brand= request.form.get('brand')
         fuelType = request.form.get('fuel') vehicletype= request.form.get('vehicletype')

        new_row("yearOfRegistration':regyear, 'powerPS':powerps, 'kilometer':kms,
        monthofRegistration': regmonth, gearbox gearbox, 'notRepairedDamage': damage,
        'model':model, 'brand':brand, 'fuelType': fuelType, 'vehicleType': vehicletype)
        print(new row)
        new_df = pd.DataFrame(columns =['vehicleType', 'yearOfRegistration', 'gearbox",
        'powerPS', 'model', 'kilometer', 'monthofRegistration', 'fuelTypek, 'brand',
        'notRepairedDamage'])

        new_df= new_df.append(new row, ignore_index= True)
         labels = ['gearbox', 'notRepairedDamage', 'model', 'brand', 'fuelType', 'vehicleType']
        mapper = {}
        for i in labels:
                mapper[i] = LabelEncoder()
                mapper[i].classes_= np.load(str('classes'+i+.npy'))
               tr= mapper[i].fit_transform(new_df[i]) new_df.loc[:, i +'_labels'] = pd.Series (tr,
        index-new_df.index) labeled = new_df[ ['yearOfRegistration' ,"powerPS' 'kilometer'
        "monthOfRegistration']+[x+'_labels' for x in labels]]
X=labeled.values
        print(X)
        y_prediction=model.rand.predict(X)
         print(y_prediction)
        return render_template('resalespredict.html',ypred = 'The resale value predicted is
        {:.2f}$'.format(y_prediction[0]))



        Run the app

        If __name__ == '__main':
app.run(host='localhost',debug = True, threaded = False)
```

13.2 GitHub

https://github.com/IBM-EPBL/IBM-Project-42303-1660658809

13.3 DEMO Link


HTML Forms
Please fill the details of your car:
Reg Year
Reg Month
Power of Car in PS
Kilometer the car has to driven
Expected Budget
Gear Box Type : Manual
Automatic
Not disclosed
Your car is damaged or Repaired: Yes
No
Not disclosed
Model Type
Brand of the car
Fuel Type : cng
Petrol
Diesel
Vehicle Type