

PROJECT ON

CAR RESALE VALUE PREDICTION

Powered by IBM India

Submitted by

YAZHINI V

SINDHU BHARATHI M

VISHAL ANANTH

MANOBALA

TEAM ID: PNT2022TMID42430

CHRIST THE KING ENGINEERING COLLEGE

KARAMADAI-641104, COIMBATORE

DEPARTMENT OF

COMPUTER SCIENCE & ENGINEERING

Mentor: Mr.Vimalaithayan AP/CSE

Industrial Mentor: Miss.Swetha

Sl.No	context	Page No
1	Introduction	3
2	Literature Survey	4
3	Ideation & Proposed Solution	6
4	Requirements Analysis	10
5	Project Design	11
6	Project Planning & Scheduling	13
7	Coding and Solutioning	14
8	Testing	15
9	Result	17
10	Advantages & Disadvantages	17
11	Conclustion	18
12	Future Scope	19
13	Appendix	21

Introduction

This project “Car Resale Value Prediction” aims to build a model to predict used cars reasonable prices based on multiple aspects, including vehicle mileage, year of manufacturing, fuel consumption, transmission, road tax, fuel type, and engine size. This model can benefit sellers, buyers, and car manufacturers in the used cars market. Upon completion, it can output a relatively accurate price prediction based on the information that users input. The model building process involves machine learning and data science. The data set used was scraped from

listings of used cars. Various regression methods, including linear regression, polynomial regression, support vector regression, decision tree regression, and random forest regression, were applied in their search to achieve the highest accuracy. Before the actual start of model-building, this project visualised the data to understand the data set better. The dataset was divided and modified to fit the regression, thus ensuring the performance of the regression.

1.1 Project Overview

A car price prediction has been a high interest research area, as it requires noticeable effort and knowledge of the field expert. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars, the applied three machine learning techniques are random forest, KNN and linear regression algorithm. Respective performances of different algorithms were then compared to find one that best suits the available data set. This ability to capture data, analyse it and use it to personalise a shopping experience or implement in the future of retail.

Parameters involved :

Car_name, Year, Selling_Price, Present_Price, Kms_Driven, Fuel_type,

Seller_type, Transmission, Owner and so on.

1.2 Purpose

Car makers face several challenges in the second-hand market. The depth crisis in the European Union, the general problem of overcapacity, increasing competition from Asian manufacturers, and the trend toward more eco-friendly cars are only a few factors that add to the difficulty of selling used vehicles in the second-hand market and decrease sales margins. Therefore, car makers require sophisticated decision support systems to sustain the profitability of the used car business.

2. LITERATURE SURVEY

2.1 Existing problem

The problem is defined as the optimised way to estimate insurance cost based on the manufacturer with some additional costs incurred by the Government in the form of taxes. As the existing methods for estimating the cost takes a lot of time and energy and due to the increased price of new cars and the inability of customers to buy new cars due to the lack of funds, used cars sales are on a global increase.

The prices of new cars in the industry is fixed by the So, customers buying a new car can be assured of the money they invest to be worthy. There is a need for a used car price prediction system to effectively determine the worthiness of the car using a variety of features. Even though there are web sites that offer this service, their prediction method may not be the best. Besides, different models and systems may contribute to predicting power for a used car's actual market value. It is important to know their actual market value while both buying and selling.

2.2 References

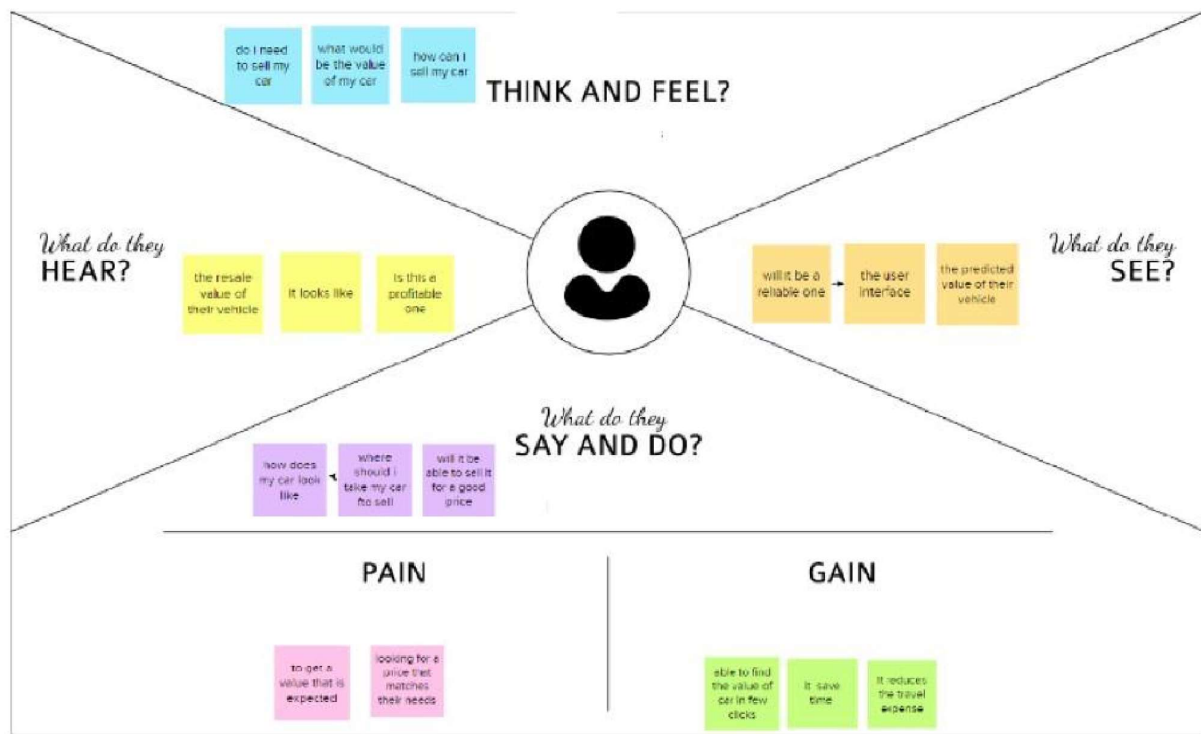
At present, under the guidance of the new generation of information technology, the rapid accumulation of data, the continuous improvement of computing power, the continuous optimization of algorithm models, and the rapid rise of multi-scene applications have made profound changes in the development environment of Machine Learning.

2.3 Problem statement definition



Ideation & Proposed Solution

Empathy Map Canvas



Ideation & Brainstorming

STEP 1:

- o Prediction using Car image. By using the exterior and interior image of the car.
- o The value will be predicted based on the appearance of the car. If there any damage or n numbers scratches the car resale value will be quite affected.
- o By using neural network value of the car can be predicted.
- o Neural network algorithm is developed by considering the human brain that takes a set of units as input and transfers results to a predefined output.

STEP 2:

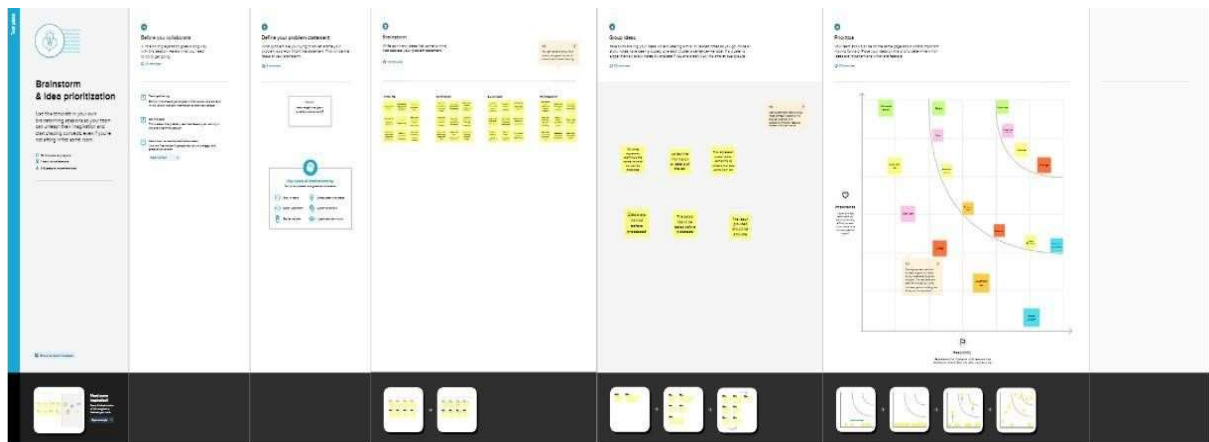
- o The main objective of this project is to predict the Prices of used cars, compare the prices and also estimate the lifespan of a particular car. o Insurance, Company claims,etc. o regression Algorithm is used to predict the value.
- o Regression model based on k-nearest neighbour machine learning algorithm was used to predict the price of a car.

STEP 3:

- o Prediction using engine car condition.
- o User should upload engine sound in the format of audio file.
- o By using Convolutional Neural Networks methodology price can be predicted.
- o CNNs for Machine Learning on sound data by spectrogram approach that was just converts each song (or song segment) into a spectrogram: a two dimensional matrix

STEP 4:

- o Economic Conditions.
- o Kilo-metres Covered.
- o Its mileage (the number of kilometres it has run) and its horsepower
- o Car prediction using XGBoost algorithm accurate results will be monitored.
- o XGBoost as a regression model gave the best M SLE and RM SE values.



3.1 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Car makers face several challenges in the second-hand market. The depth crisis in the European Union, the general problem of overcapacity, increasing competition from Asian manufacturers, and the trend toward more eco-friendly cars are only a few factors

		that add to the difficulty of selling used vehicles in the second-hand market and decrease sales margins. Therefore, car makers require sophisticated decision support systems to sustain the profitability of the used car business.
2.	Idea / Solution description	The main aim of this project is to predict the price of used cars using the various Machine Learning(ML) models. This can enable the customers to make decisions based on different inputs or factors namely, Brand or Type of the car, Model of the car, Location, Year of manufacturing, Type of fuel, Price range or Budget, Mileage to name a few characteristic features required by the customer. The project Car Resale Value Prediction deals with providing the solution to these problems. Through this project, we will get to know which of the factors are significant and tell us how they affect the car's worth in the market.
3.	Novelty / Uniqueness	Deciding whether a used car is worth the posted price when you see listings online can be difficult. Several factors, including mileage, make, model, year, etc. can influence the actual worth of a car.
4.	Social Impact / Customer Satisfaction	Customers are highly satisfied with high resale value. In this prediction is used to customer can easily find and buy the car he was looking forward to.
5.	Business Model (Revenue Model)	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
6.	Scalability of the Solution	The implementation of our solution helps to find the accurate value of the used care. 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. We will be using various regression algorithms and algorithm with the best accuracy will be taken as a solution,

Proposed Solution fit

Define CS, fit into CC	<div>1. CUSTOMER SEGMENT(S) Who is your customer? (e.g. working professionals 25 yrs. and)</div> <div>Used car sellers</div>	<div>6. CUSTOMER CONSTRAINTS What constraints prevent your customers from taking action or limit their choices of options? (e.g. spending power, budget, no cash, internet connection, mobile device)</div> <div><ul style="list-style-type: none">To determine the worthiness of the car by their own within few minutesA loss function is to be optimized by spending money for dealers, brokers to buy or sell a car.</div>	<div>8. AVAILABLE SOLUTIONS Which solutions are available to the customer when they face the problem? (e.g. Which ones to consider these solutions how? or how one is superior or inferior to the original existing products getting job done? Where else they have to be part?)</div> <div><ul style="list-style-type: none">In the past User cannot find the value of used car buy their own without prior knowledge about cars.A person who don't know much about the car can also make predictions for used cars easily.</div>	Explore AS, differentiate
	<div>2. JOBS TO BE DONE / PROBLEMS Which jobs to be done (or problems) do you address for your customer? These could be more than one, explore different roles.</div> <div>To build a supervised machine learning model using regression algorithms for forecasting the value of a vehicle based on multiple attributes such as</div> <div><ul style="list-style-type: none">Condition of EngineYear of RegistrationKilometersNumber of Owner</div>	<div>9. PROBLEM ROOT CAUSE What is the real reason that the problem exists? What is the link more behind the need to do this job? (e.g. customers have to do it because of the change in regulations)</div> <div><ul style="list-style-type: none">The price predicted by the dealers or brokers for used car is not trustfulUsers can predict the correct valuation of the car remotely without human intervention like car dealers.User can eliminate the valuation predicted by the dealer</div>	<div>7. BEHAVIOUR What does your customer do to address the problem and see the solution? (e.g. identify online, find the right value panel from the, calculate stage and benefits, identify, understand, customer spend time on understanding, make (e.g. Google search)</div> <div><ul style="list-style-type: none">The History of Your Car's condition and documents produced by them will be Suspicious.The model is to be built would give the nearest value of the vehicle by eliminating anonymous value predicted by using humans.</div>	Focus on J&P, map into BE, understand RC
Focus on J&P, map into BE, understand RC	<div>3. TRIGGERS What trigger customers to act? (e.g. comparing their neighbours, installing sales panel, website about a more efficient solution in the cars)</div> <div>Users can predict the correct valuation of the car by their own like Olxcars, Cars24 and other car resale value prediction websites by using model, year, owner, etc.</div>	<div>10. YOUR SOLUTION If you are working about creating business, you'd have your current solution that did in the market, and check how much it is working. If you are doing only one business proposition, then keep it that and you'll in the customer and strategy with a solution that do value without foundation, solve a problem and maintain customer behaviour.</div> <div><ul style="list-style-type: none">The main aim of this project is to predict the price of used cars using the Machine Learning (ML) algorithms and collection data's about different cars.</div>	<div>8. CHANNELS of BEHAVIOUR K1 ONLINE What kind of channels do customers take online? (Search online channels from #) K2 OFFLINE What kind of channels do customers take offline? (Search offline channels from #) and use them for customer development.</div> <div><ul style="list-style-type: none">Customer should predict the worth of the car by using different parameters given by the owner.</div>	
	<div>4. EMOTIONS: BEFORE / AFTER How do customers feel when they face a problem or a job and afterwards? (e.g. how customer's confidence, is positive, can it be your communication strategy & design)</div> <div>Before:<ul style="list-style-type: none">User will be in fear about the biased values predicted by the humans based on the condition of the car.After:<ul style="list-style-type: none">User can determine the worthiness of the car by their own without human intervention.</div>	<div>The project should take parameters related to used car as inputs and enable the customers to make decisions by their own.</div>	<div><ul style="list-style-type: none">User Should confirm the details provided about the vehicle in RTO online.User can decide by seeing the exterior and interior condition of the car.User can test the performance of the car and to buy it up in a affordable price based on its condition.</div>	
Identify strong TR & EM				

REQUIREMENT ANALYSIS

Functional Requirement

The functional Requirements of this projects involves the better understanding of

Pre-processing, Application designing using HTML & CSS and

IBM Watson Cloud. IBM Watson provides the services such as Database, deployment etc.

Hardware requirements Operating system- Windows 7,8,10

Software Requirements

- Python
- VS code
- PIP 2.7
- Jupyter Notebook
- Chrome

Non Functional Requirement

The Non - Functional Requirements of this project are,

- Highly accurate Image Predictive model
- better user responsive web application
- Cloud database for storing the informations

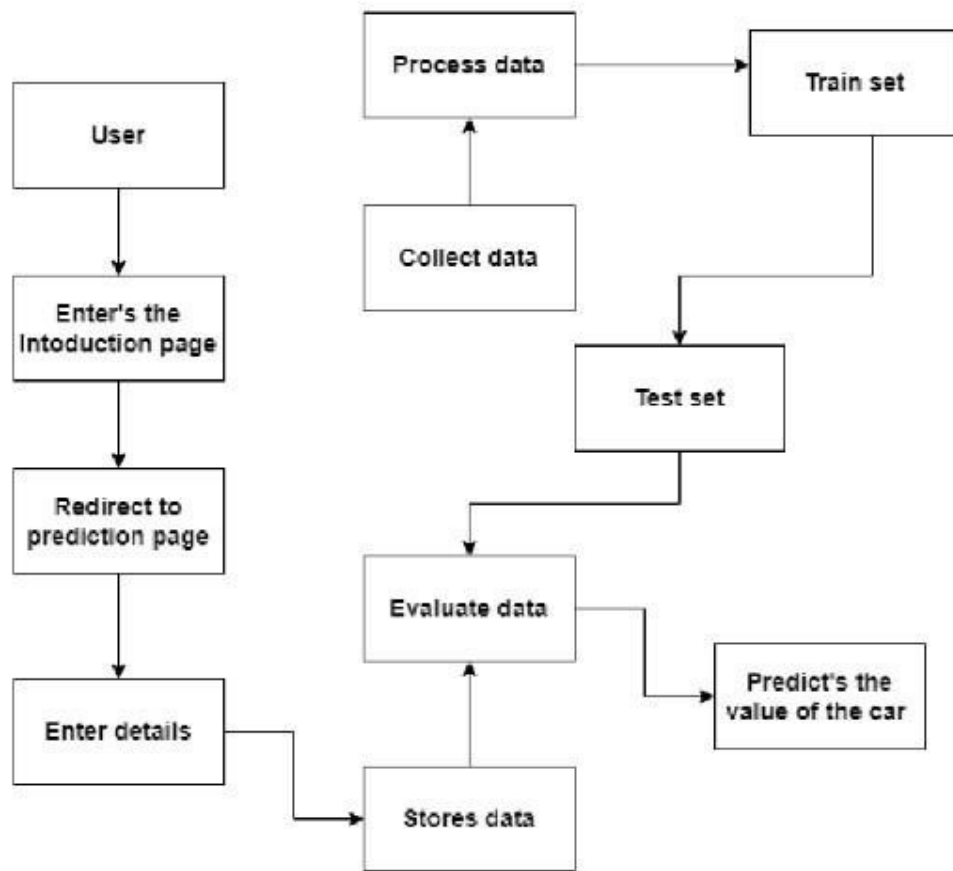
PROJECT DESIGN

5.1 Data Flow Diagrams

5.2 Solution & Technical Architecture

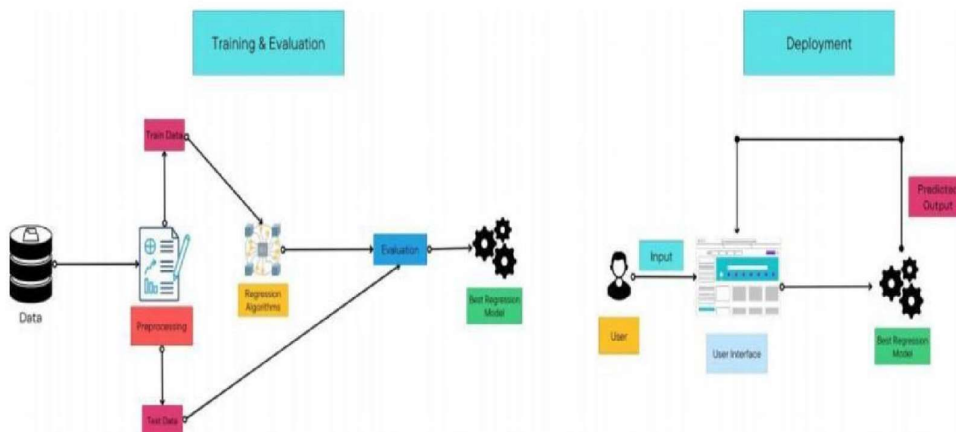
5.2 Solution & Technical Architecture

5.2 Solution & Technical Architecture



5.2 Solution & Technical Architecture

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Pre-process data	USN-1	Collect Dataset	1	Low	Yazhini V
Sprint-1		USN-2	Import required libraries	1	Low	Sindhu Bharathi M
		USN-3	Read and clean data sets	2	Low	Vishal Ananth R
Sprint-2	Model building	USN-1	Split data into independent and dependent variables	3	Medium	Manobala F
Sprint-2		USN-2	Apply using regression model	3	Medium	Yazhini V
Sprint-3	Application building	USN-1	Build python flask application and HTML page	5	High	Sindhu Bharathi M
Spint-3		USN-2	Execute and test	5	High	Manobala F & Vishal Ananth R
Spint-4	Training the model	USN-1	Train machine learning model	5	High	Sindhu Bharathi M
Spint-4		USN-2	Integrate flask	5	High	Sindhu Bharathi M & Vishal Ananth



5.3 User Stories

PROJECT PLANNING

6.1 Sprint Planning and Estimation

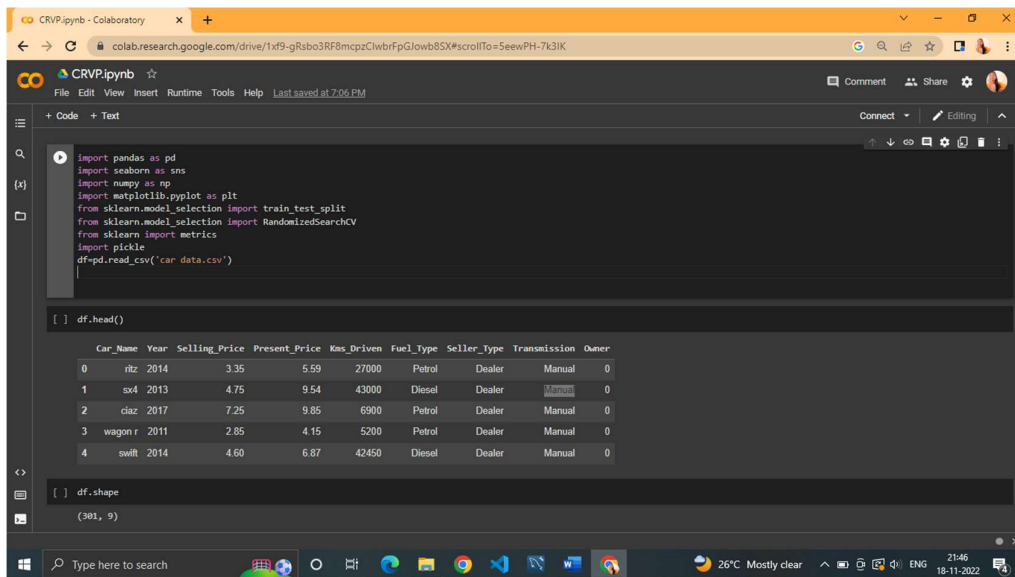
6.2 Sprint Delivery Schedule

- Pre -requisites
- Import Required libraries
- Collect Data Set
- Pre the process the data
- Choose the Appropriate Model
- Train the model on IBM
- Integrate with Flask endpoint
- Index..html
- Registration form.html
- Flask application
- App.py

7. CODING & SOLUTION

Feature 1

7.2 Feature 2



The screenshot shows a Google Colab notebook with the following code and output:

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.model_selection import RandomizedSearchCV
from sklearn import metrics
import pickle
df=pd.read_csv('car data.csv')
```

Below the code, the command `df.head()` is executed, resulting in a preview of the first five rows of the 'car data.csv' file:

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	claz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0

At the bottom, the command `df.shape` is executed, showing the dimensions of the dataset: `(381, 9)`.

```
from flask import Flask, render_template, request
import pickle
import numpy as np
from sklearn.preprocessing import StandardScaler
app = Flask(__name__)
model = pickle.load(open('random_forest_regression_model.pkl', 'rb'))
@app.route('/', methods=['GET'])
def Home():
    return render_template('base.html')

standard_to = StandardScaler()

@app.route('/predict_page')
def predict_page():#predicting page
    return render_template('index1.html')
@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'])
        Kms_Driven=int(request.form['Kms_Driven'])
        Kms_Driven2=np.log(Kms_Driven)
        Owner=int(request.form['Owner'])
        Fuel_Type_Petrol=request.form['Fuel_Type_Petrol']
        if(Fuel_Type_Petrol=='Petrol'):
            Fuel_Type_Petrol=1
            Fuel_Type_Diesel=0
        else:
            Fuel_Type_Petrol=0
            Fuel_Type_Diesel=1
        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_Mannual=request.form['Transmission_Mannual']
if(Transmission_Mannual=='Mannual'):
    Transmission_Mannual=1
else:
    Transmission_Mannual=0
prediction=model.predict([[Present_Price,Kms_Driven2,Owner,Year,Fuel_Type_
Diesel,Fuel_Type_Petrol,Seller_Type_Individual,Transmission_Mannual]])
output=round(prediction[0],2)
if output<0:
    return render_template('index1.html',prediction_texts="Sorry you
cannot sell this car")
else:
    return render_template('index1.html',prediction_text=" (You Can Sell
the Car at){}".format(output))
else:
    return render_template('index1.html')

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

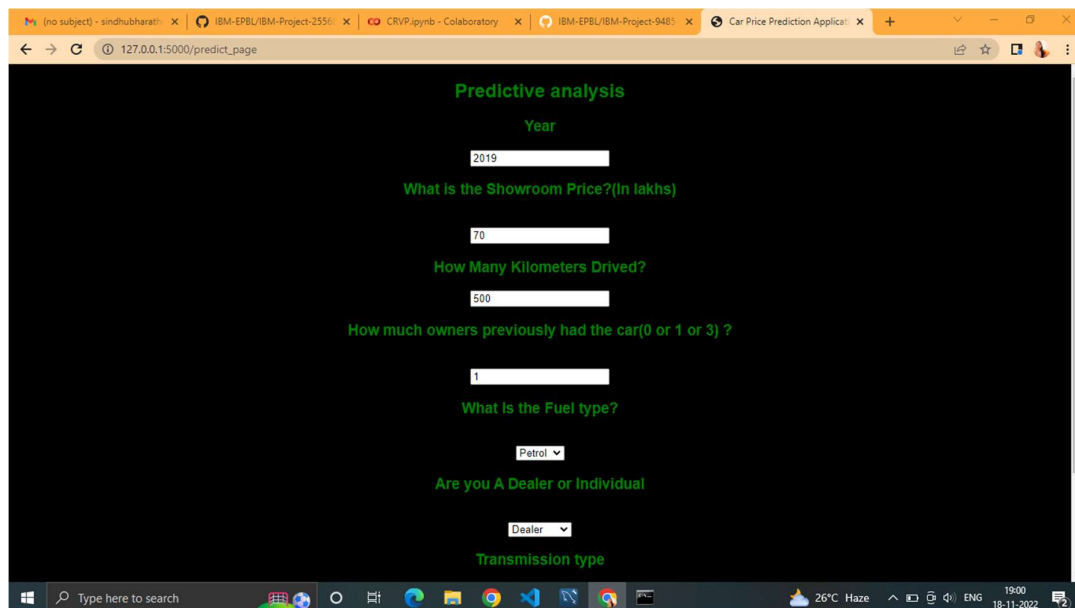
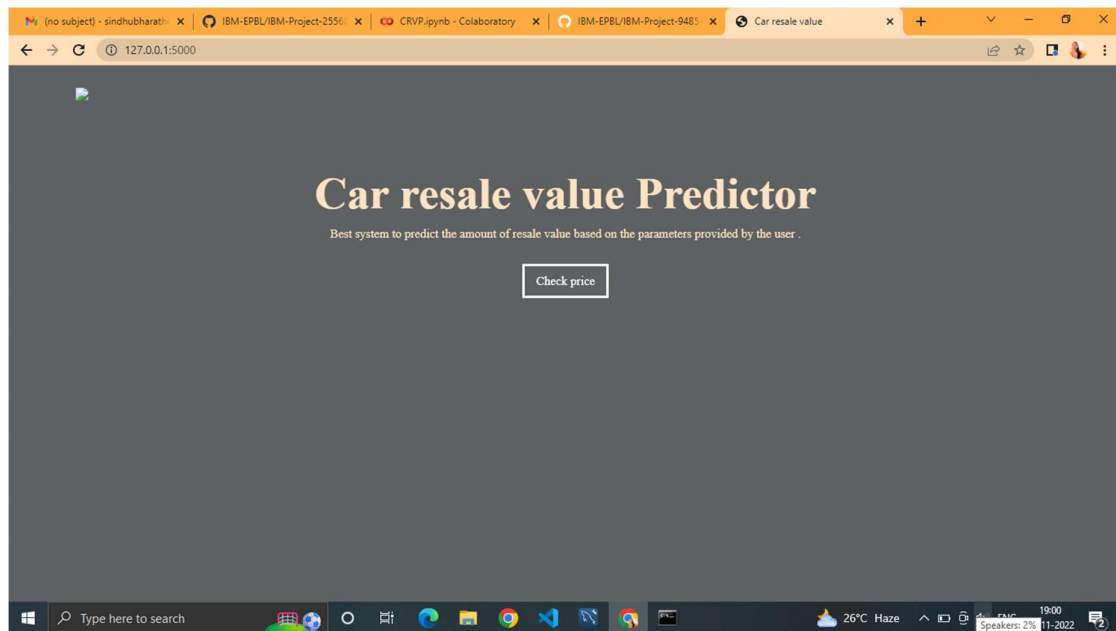
```

8. TESTING

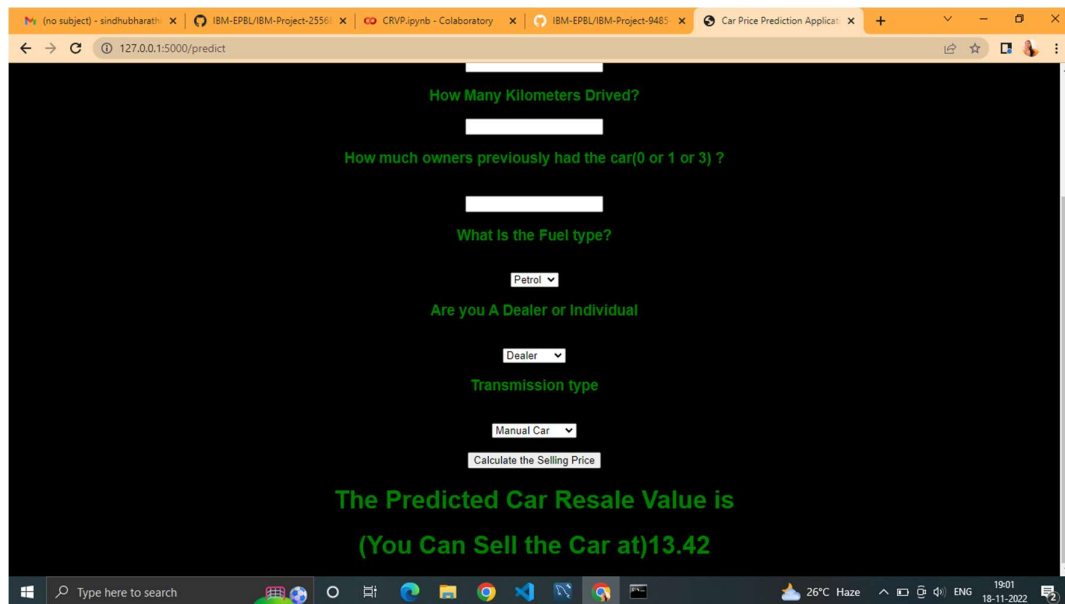
8.1. Test Cases

1. User Login and Registration test
2. Database Update test
3. Prediction test

8.2. User Acceptance Testing



The login web page is tested with the invalid user information to check the invalid login testing into the webpage.



9. PERFORMANCE

9.1 Performance metrics

```
{'mae': 1325.112086905962,  
'mse': 9577053.62710202,  
'rmse': 3094.6815065692977,  
'rmsle': 8.03744027403009,  
'r2': 0.8661221626879432,  
'adj_r2_score': 0.8661152969113608}
```

The model is tested with the various damaged car images which is not used during the training and validation of the model which also shows that the model works with the accuracy of about 98% in the overall performance

10. ADVANTAGES AND DISADVANTAGES

- To develop an efficient and effective model which predicts the price of a used car according to the user's inputs and achieve good accuracy.

CONS:

- Less effective

11. CONCLUSION

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.

12.FUTURE SCOPE

In future this machine learning model may bind with various websites 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 a user interface for interacting with users. 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 dataset.

Github Repo: <https://github.com/IBM-EPBL/IBM-Project-7348-1658853071>

APPENDIX

App.py

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

```

from sklearn.preprocessing import StandardScaler
app = Flask(__name__)
model = pickle.load(open('random_forest_regression_model.pkl', 'rb'))
@app.route('/',methods=['GET'])
def Home():
    return render_template('base.html')

standard_to = StandardScaler()

@app.route('/predict_page')
def predict_page():#predicting page
    return render_template('index1.html')
@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'])
        Kms_Driven=int(request.form['Kms_Driven'])
        Kms_Driven2=np.log(Kms_Driven)
        Owner=int(request.form['Owner'])
        Fuel_Type_Petrol=request.form['Fuel_Type_Petrol']
        if(Fuel_Type_Petrol=='Petrol'):
            Fuel_Type_Petrol=1
            Fuel_Type_Diesel=0
        else:
            Fuel_Type_Petrol=0
            Fuel_Type_Diesel=1
        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_Mannual=request.form['Transmission_Mannual']
        if(Transmission_Mannual=='Mannual'):
            Transmission_Mannual=1
        else:
            Transmission_Mannual=0
        prediction=model.predict([[Present_Price,Kms_Driven2,Owner,Year,Fuel_Type_Diesel,
Fuel_Type_Petrol,Seller_Type_Individual,Transmission_Mannual]])
        output=round(prediction[0],2)
        if output<0:
            return render_template('index1.html',prediction_texts="Sorry you cannot sell
this car")
        else:
            return render_template('index1.html',prediction_text=" (You Can Sell the Car
at){}".format(output))
    else:
        return render_template('index1.html')

if __name__=="__main__":

```

```
app.run(debug=True)
```