PROJECT ON

CAR RESALE VALUE PREDICTION

Powered by IBM India

Submitted by

ARUN KUMAR G

DHARSHITH K

HARI KUMAR G

KAVINESH T

TEAM ID: PNT2022TMID19654

NANDHA ENGINEERING COLLEGE

VAIKKALMEDU , ERODE - 638052

DEPARTMENT OF

ELECTRONICS AND
COMMUNICATION ENGINEERING

Mentor: Mr.Jayachandran T(AP/ECE)

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 a spects, including vehicle mileage, year of manufacturing, fuel consumption, transmission, road tax, fuel type, and engine size. This model can benefit seller s, buyers, and car manufacturer s in the used car s 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 there search to achieve the highest accuracy. Before the actual start of model-building, this project visualised the data to under stand 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, a s it require s noticeable effort and knowledge of the field expert. Considerable number of distinct attribute s are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars, the applied three machine learning technique s are random forest ,KN-N and linear regression algorithm. Respective performance s of different algorithm s were then compared to find one that be stsuits the available data set. This ability to capture data, analyse it and use it to

personalise a shopping experience or implement is 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

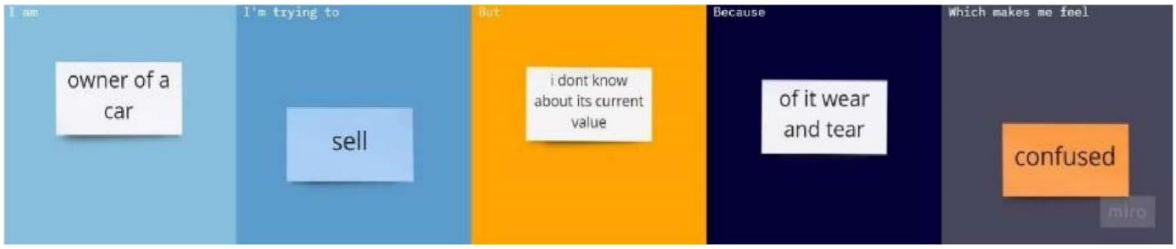
The problem is defined a s the optimised way to estimate insurance cost based on the manufacturer with some additional costs incurred by the Government in the form of taxes. A s the existing method s for estimating the cost takes a lot of time and energy and due to the increased price of new cars and 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 site s that offer this service, their prediction method may not be the best. Beside s, 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:

- Prediction using Car image. By using the exterior and interior image of the car.
- 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.
- By using neural network value of the car can be predicted.
- 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:

- 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.
- Regression model based on k-nearest neighbour machine learning algorithm was used to predict the price of a car.

STEP 3:

- Prediction using engine car condition.
- User should upload engine sound in the format of audio file.
- By using Convolutional Neural Networks methodology price can be

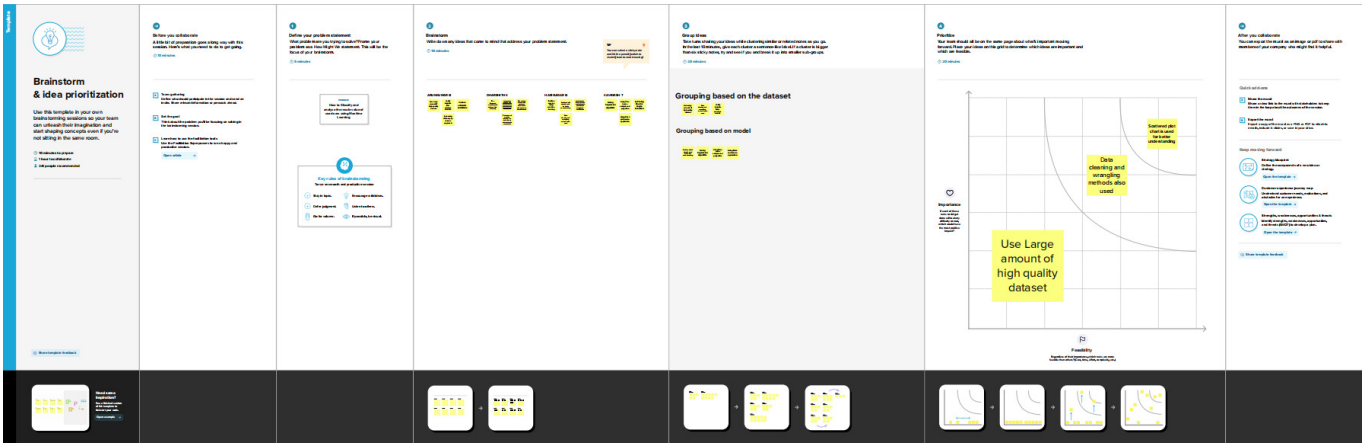
predicted.

- 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.



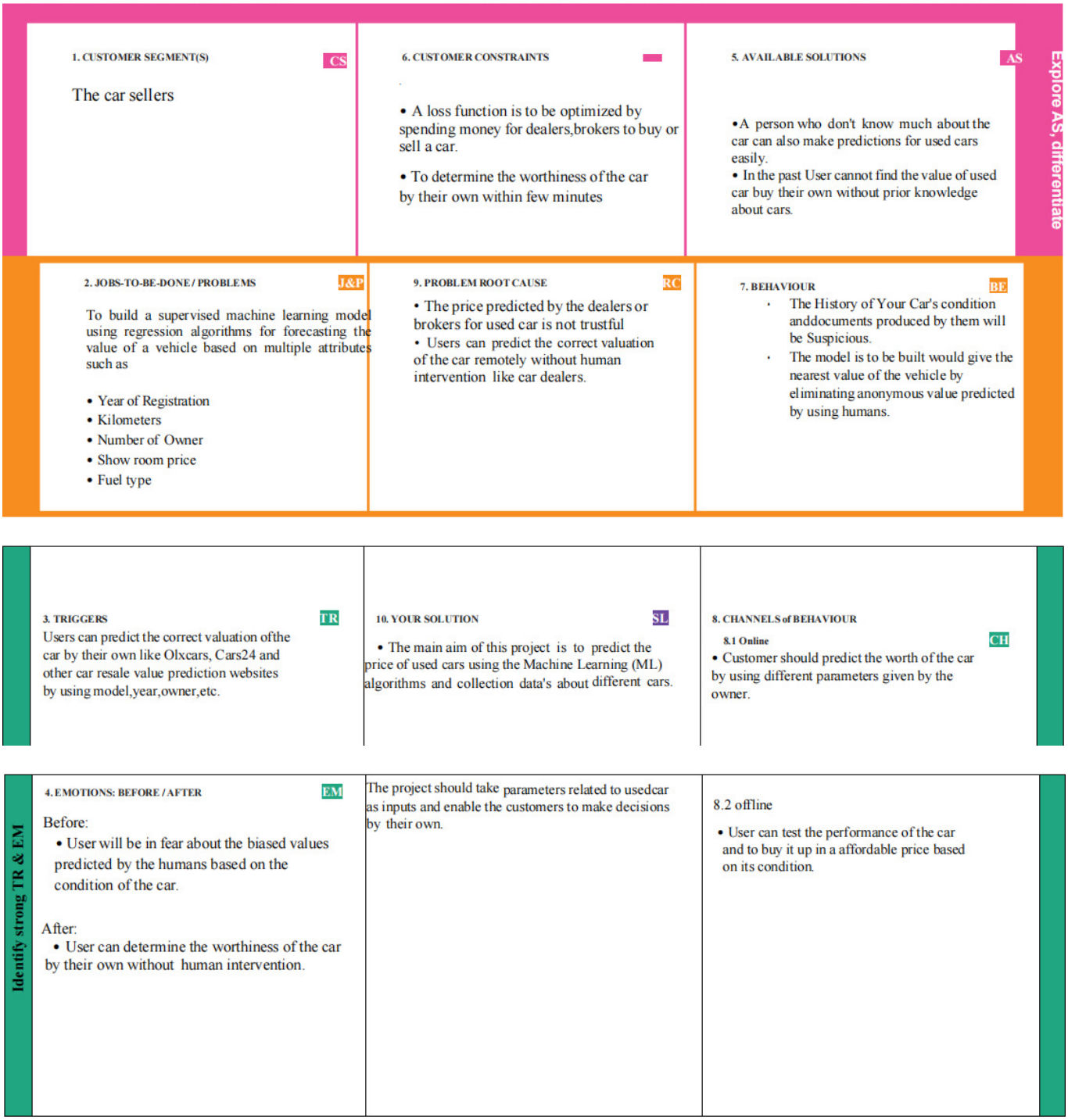
3.1 Proposed Solution

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

Project Title: Car Resale Value Prediction | **Project Design Phase-I -Problem Solution Fit** | **Team ID:** PNT2022TMID16869



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

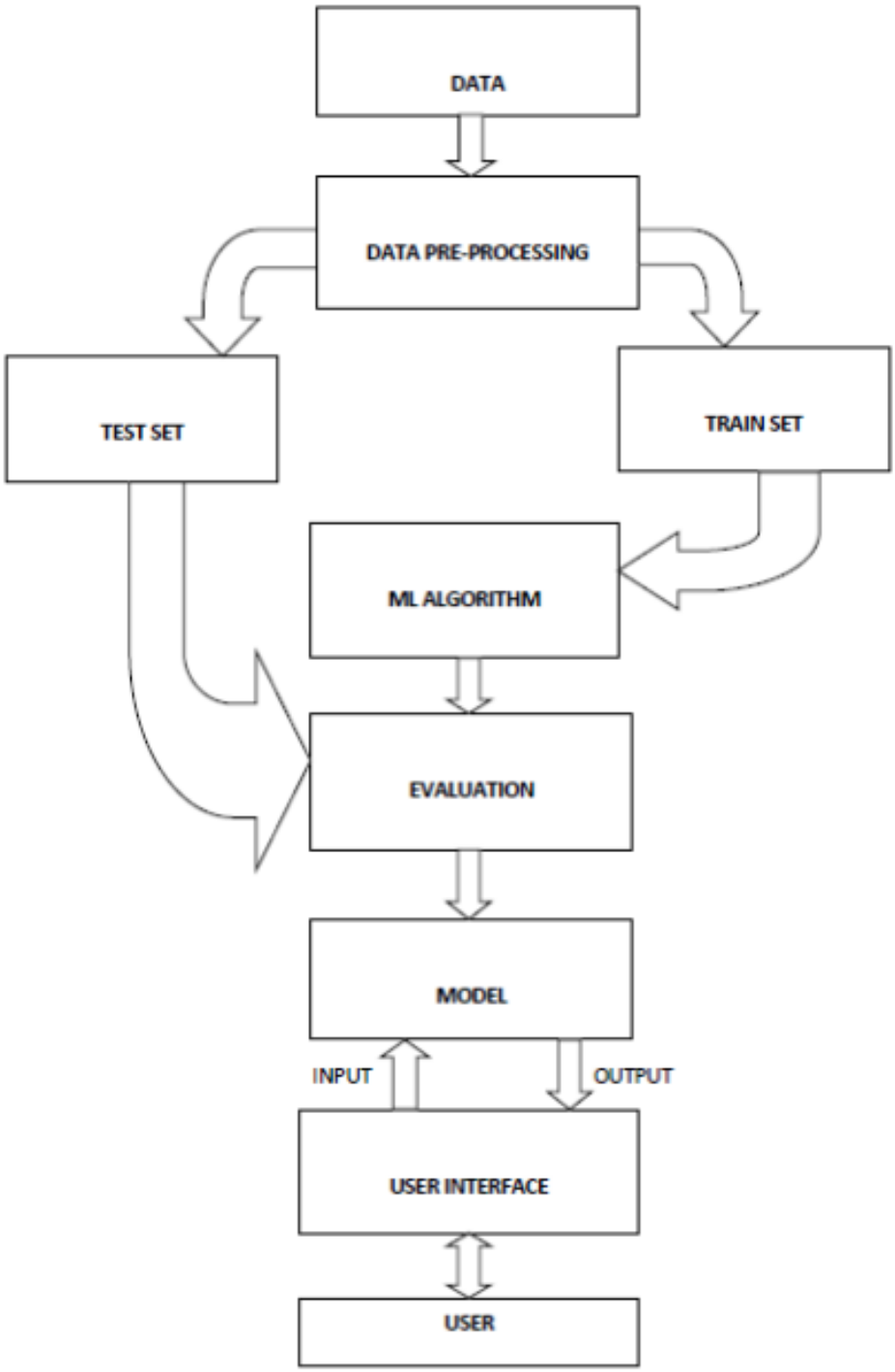
PROJECT DESIGN

5.1 Data Flow Diagrams

5.2 Solution & Technical Architecture

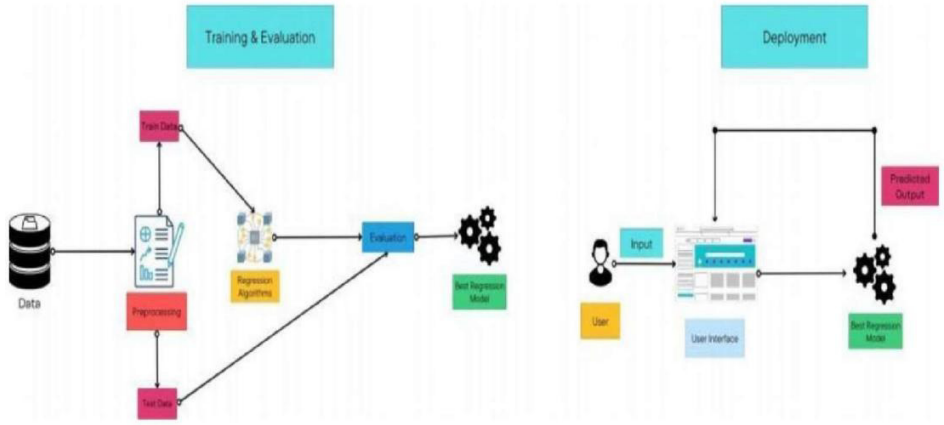
5.2 Solution & Technical Architecture

5.2 Solution & Technical Architecture


DATA
DATA PRE-PROCESSING
TEST SET
TRAIN SET
ML ALGORITHM
EVALUATION
MODEL
INPUT
OUTPUT
USER INTERFACE
USER

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	Hari Kumar G
Sprint-1		USN-2	Import required libraries	1	Low	Kavinesh T
		USN-3	Read and clean data sets	2	Low	Dharshith K
Sprint-2	Model building	USN-1	Split data into independent and dependent variables	3	Medium	Arun kumar G
Sprint-2		USN-2	Apply using regression model	3	Medium	Arun kumar G
Sprint-3	Application building	USN-1	Build python flask application and HTML page	5	High	Kavinesh T
Spint-3		USN-2	Execute and test	5	High	Dharshith K &Kavinesh T
Spint-4	Training the model	USN-1	Train machine learning model	5	High	Arun Kumar G
Spint-4		USN-2	Integrate flask	5	High	Arun Kumar G & Harikumar G



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
- Regi stration form.html
- Fla sk application
- App.py

7. CODING & SOLUTION

Feature 1

7.2 Feature 2

Application Building:

Build The Python Flask App

#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)
```

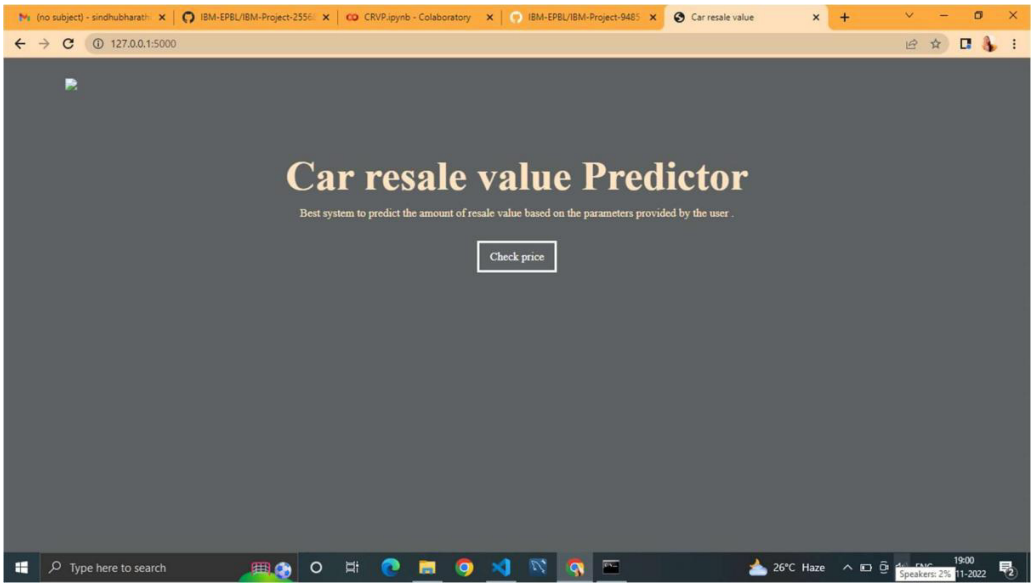
```
y_prediction=model.rand.predict(X)
print(y_prediction)
return render_template('resalespredict.html',ypred =  'The resale value predicted is
{:.2f}$'.format(y_prediction[0]))


Run the app

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

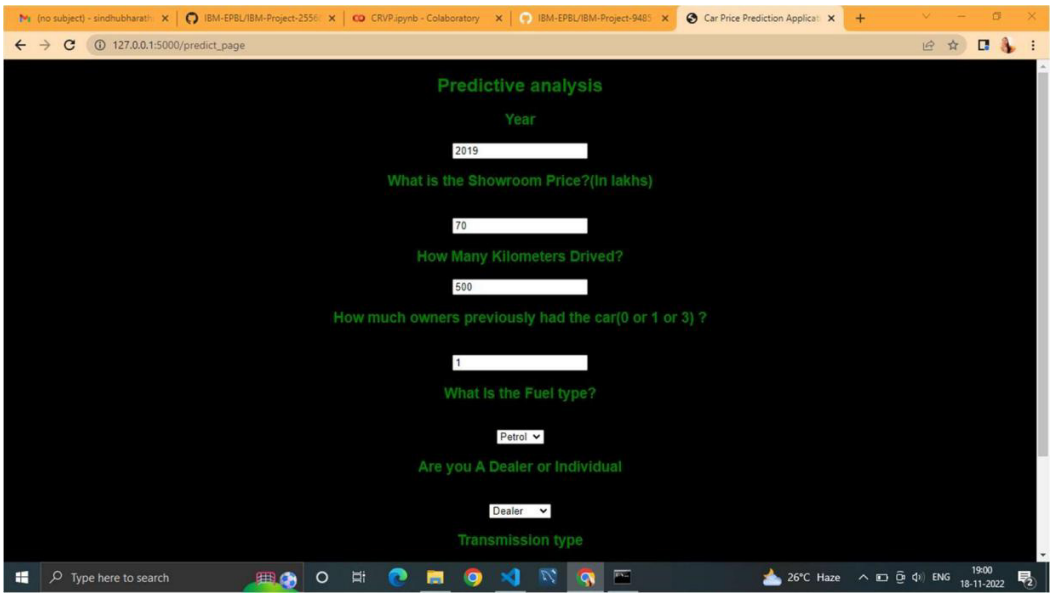
8. TESTING

8.1. Test Cases

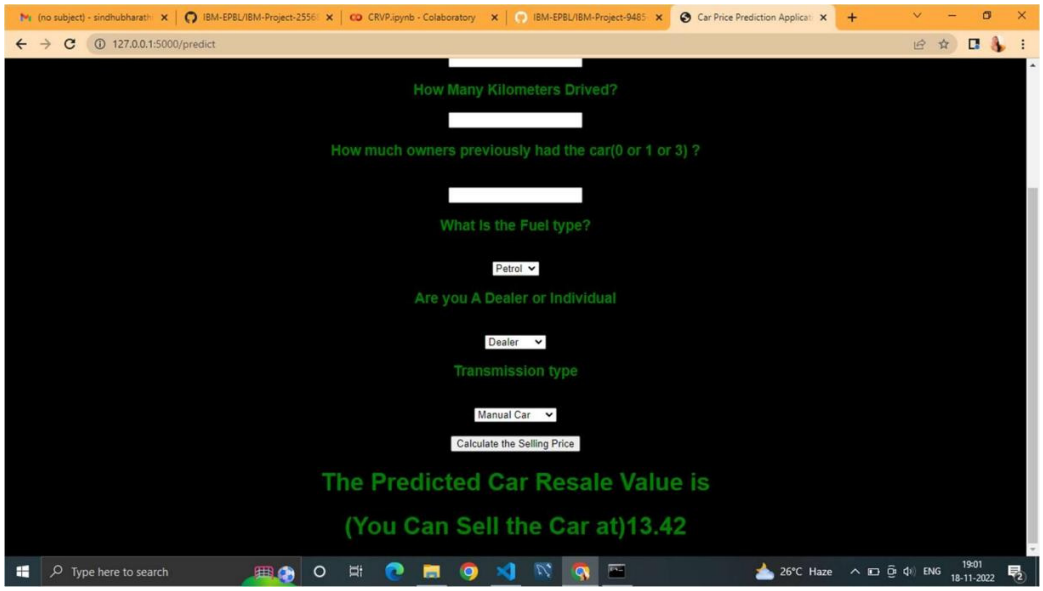
1. User Login and Registration test
2. Database Update test
3. Prediction test

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

Application Building:

Build The Python Flask App

#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)
```