Project On

# Car Resale Value Prediction

### Powered by IBM India

**Submitted by**

Ajay S

Anandha surya MS

Senthilkamalesh M S C

Siva Vignesh M

**PROJECT ID**: **PNT2022TMID23218**

Of

## Department of Information Technology

**Velammal college of Engineering and Technology , Madurai - 625201**

Mentor: Jegadeesan S

Industrial Mentor: Prof Swetha

**SI. No. Context Page No.**

1. [Introduction 2](#_TOC_250006)
2. [Literature Survey 4](#_TOC_250005)
3. Ideation and Proposed Solution 6
4. Requirements Analysis 10
5. [Project Design 11](#_TOC_250004)
6. Project Planning & Scheduling 13
7. Coding and Solutioning 14
8. [Testing 16](#_TOC_250003)
9. Result 18
10. Advantages & Disadvantages 19
11. [Conclusion 20](#_TOC_250002)
12. [Future Scope 21](#_TOC_250001)
13. [Appendix 22](#_TOC_250000)

# 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 user’s input. The model building process involves machine learning and data science. The dataset 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 the research to achieve the highest accuracy. Before the actual start of model-building, this project visualized the data to understand the dataset better. The dataset was divided and modified to fit the regression, thus ensuring the performance of the regression.

* 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, KN-N 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, analyze it and use it to personalize a shopping experience or implement is the future of retail.

Parameters involved:

Vehicle name; Year; Selling Price; Present Price; Kms Driven; Fuel type. Seller type; Transmission; Owner and so on.

* 1. **Purpose**

Car resale value prediction helps the user to predict the re sale value of the car

depending upon various features like kilometers driven, fuel type, etc. This resale value prediction system is made for general purpose to just predict the amount that can be roughly acquired by the user. The most essential elements for forecast are brand and model, period use of vehicle, mileage of vehicle, gear type and fuel type utilized in the vehicle just as fuel utilization per mile profoundly influences cost of a vehicle because of continuous changes in the cost of a fuel. In view of the differing highlights and factors, and furthermore with the assistance of master information the vehicle value forecast has been done precisely

# LITERATURE SURVEY

problem is defined as the optimized way to estimate in surance 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 are 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 websites 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.

## References

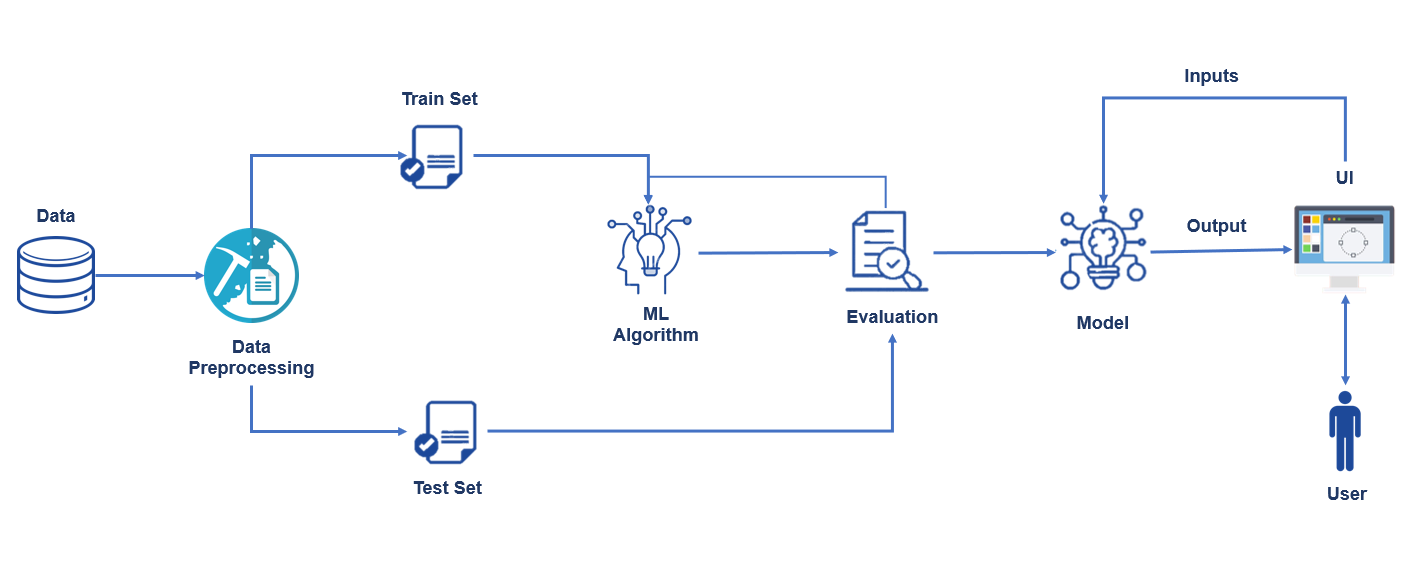
At present, under the guidance of the new generation of information technology, the rapid accumulation of data, the continuous improvement of computing power, the continuous optimization of algorithm models, and the rapid rise of multi-scene applications have made profound changes in the development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

* + 1. **Kanwal Noor, 2017, Vehicle Price Prediction System using Machine Learning Techniques International Journal of Computer Applications. Volume 167 - Number 9**
    2. **Mariana Lusitania et al, (2009). Support vector regression analysis for price prediction in a vehicle leasing application [3] Richardson, M. S. (2009). Determinants of used vehicle resale value.**
    3. **Listiani, M. (2009). Support vector regression analysis for price prediction in a car leasing application (Doctoral dissertation, Master thesis, TU Hamburg-Harburg).**
    4. **T. D. Phan, "Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City Australia", *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, pp. 35-42, 2018.**
    5. **K. Samruddhi and R. Ashok Kumar, "Used Car Price Prediction using K-Nearest Neighbor Based Model", *International Journal of Innovative Research in Applied Sciences and Engineering*, vol. 4, no. 3, pp. 686-689, 2020.**
    6. **O. Celik and U. O. Osmanoglu, "Prediction of The Prices of Second-Hand Cars", *Avrupa Bilim ve Teknoloji Dergisi*, no. 16, pp. 77-83, Aug. 2019.**
  1. **Problem statement definition**

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.

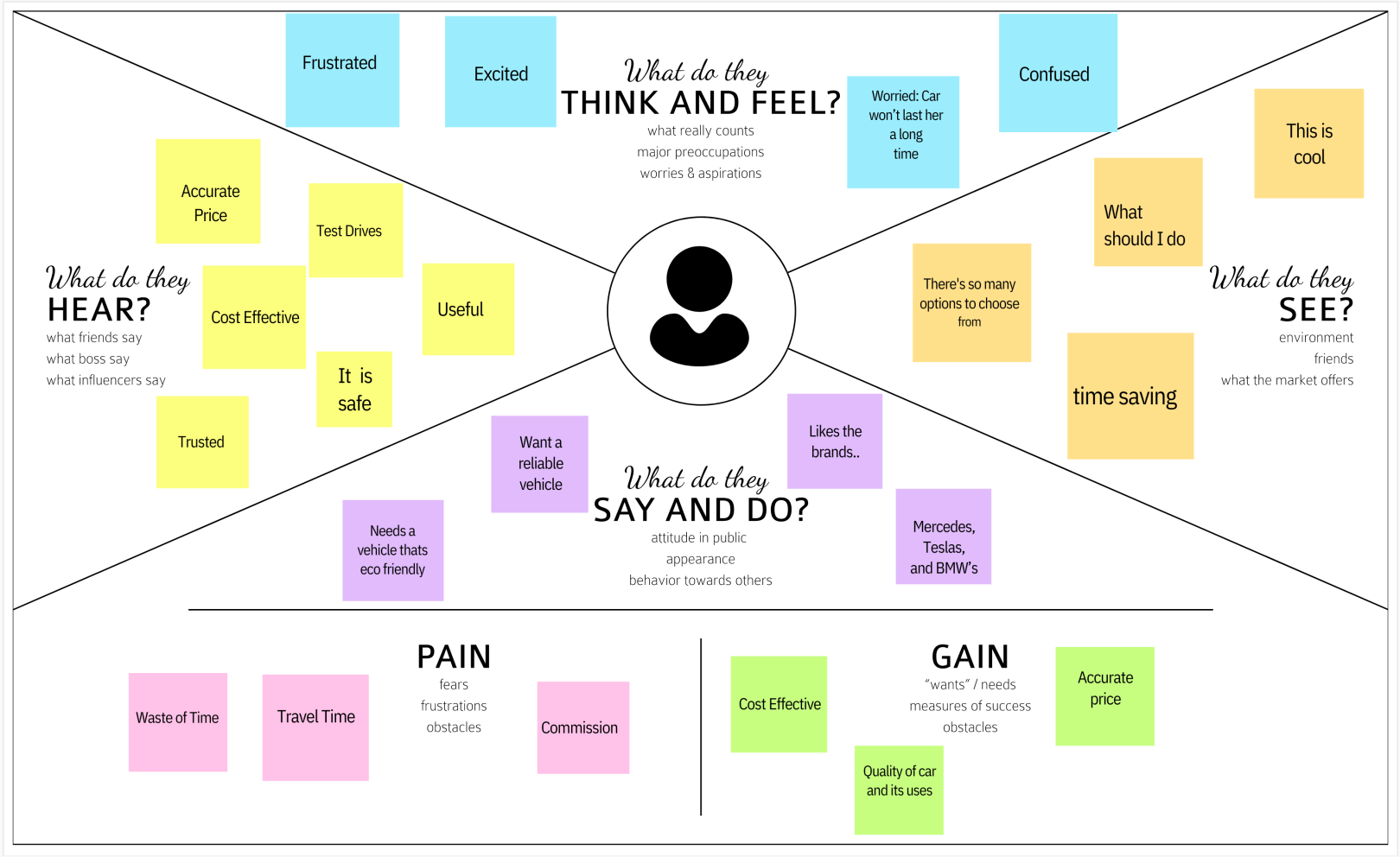
In order to predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale 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.

**Technical Architecture:**



# IDEATION & PROPOSED SOLUTION

## Empathy Map Canvas

****

* 1. **Ideation & Brainstorming**

**STEP 1:**

1. Prediction using Car image.
2. By using the exterior and interior image of the car.
3. 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.
4. By using neural network value of the car can be predicted.
5. 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:**

1. The main objective of this project is to predict the Prices of used cars, compare the prices and estimate the lifespan of a particular car.
2. Insurance, Company claims, etc.

o regression Algorithm is used to predict the value.

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

**STEP 3:**

1. Prediction using engine car condition.

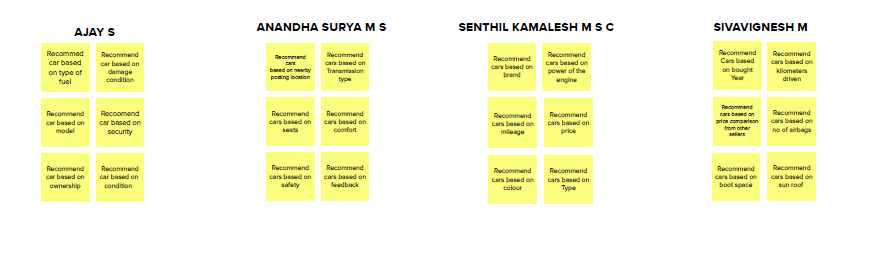
oUser should upload engine sound in the format of audio file.

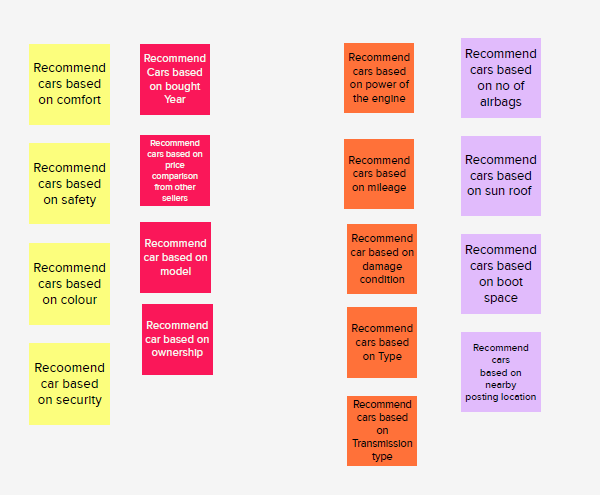
1. By using Convolutional Neural Networks methodology price can be predicted.
2. CNNs for Machine Learning on sound data by spectrogram approach that was just converts each song (or song segment) into a spectrogram: a two-dimensional matrix

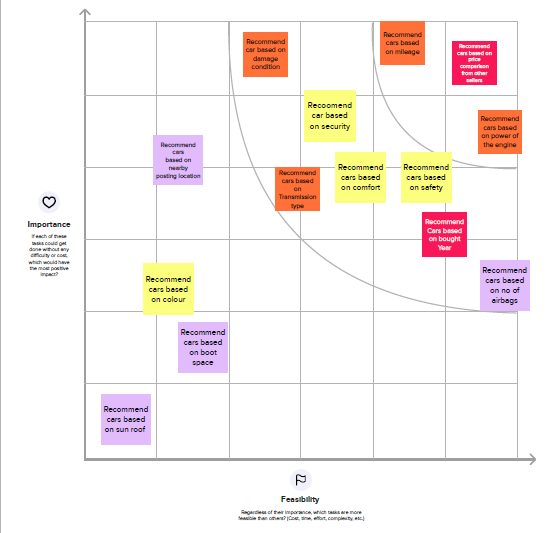
**STEP 4:**

Economic Conditions. o Kilometers Covered.

* Its mileage (the number of kilometers it has run) and its horsepower
* Car prediction using XG Boost algorithm accurate results will be monitored.
* XG Boost as a regression model gave the best MSLE and RMSE values.



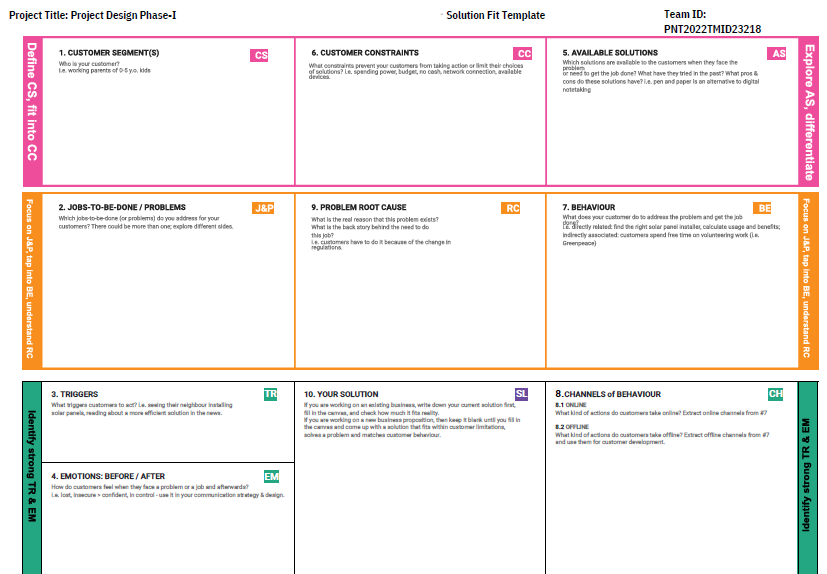




## Proposed Solution

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Parameter** | **Description** |
|  | Problem Statement (Problem to be solved) | For the purposes of car valuation, popular guides tend not to use machine learning. Instead, they source data from local sales and average the prices of many similar cars. This method works well if you have a common car with a common set of features. The condition of the car is judged very roughly, typically on a scale of one to three. Cars that are “unusual” are therefore hard to evaluate. Effectively, no inferences are drawn from similar cars but from a different make and model, whereas with machine learning, the entirety of the dataset and its features are used to train the model predictions. Using machine learning is a solution to the problem of utilization of all the data and will assist in utilizing all the features of a car to make valuations. |
|  | Idea / Solution description | New cars of a particular make, model, and year all have the same retail price, excluding optional features. This price is set by the manufacturer. Used car, however, are subject to supply-and-demand pricing. Further, used cars have additional attributes that factor into the price. These include the condition, milage, and repair history, which sets cars that may have shared a retail price apart. |
|  | Novelty /Uniqueness | The purpose of this thesis is to evaluate several different machine learning models for used car price prediction and draw conclusions about how they behave. This will deepen the knowledge of machine learning applied to car valuations and other similar price prediction problems. |
|  | Social Impact / Customer Satisfaction | This work will focus on answering the research questions. They all entail a comparison of different ML algorithms for price prediction. This will be accomplished by sourcing and preparing a dataset on which all the algorithms can be trained on and compared fairly. The algorithms selected must therefore be similar enough for the same dataset to be used for all of them. This also means that no large optimization efforts on the dataset will be made to boost the performance, if these changes do not benefit the other models. Maximizing price prediction performance of any one algorithm in ways that do not offer better comparisons is outside the scope of this work. |
|  | Business Model (Revenue Model) | A revenue model is a blueprint that shows how a startup business will earn revenue or gross income from its standard business operations, and how it will pay for operating costs and expenses |
|  | Scalability of the Solution | which of the models and parameters gives the best overall accuracy in making price predictions for used cars. The optimal parameters were determined in the process of implementing the models, and thus each model was implemented with the parameters that yielded the best performance by trial and error  All of the models approximated geometric appreciation, meaning that a constant percentage of value is lost every year independent of the age of the vehicle. Random Forest Regression had a significantly higher assessed average depreciation at approximately 13.8%, compared to the others with 9.7%. This is closer to the range of 15%-31% assessed by Karl Storchmann in his analysis of international depreciation rates |

* 1. **Proposed Solution fit**

****

# 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
    - Visual Studio Code
    - PIP 2.7
    - Jupyter Notebook
    - Chrome

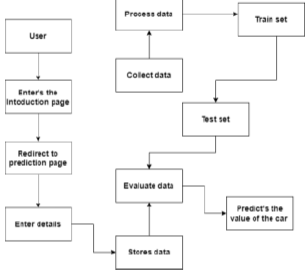
## Non-Functional Requirement

The Non - Functional Requirements of this project are,

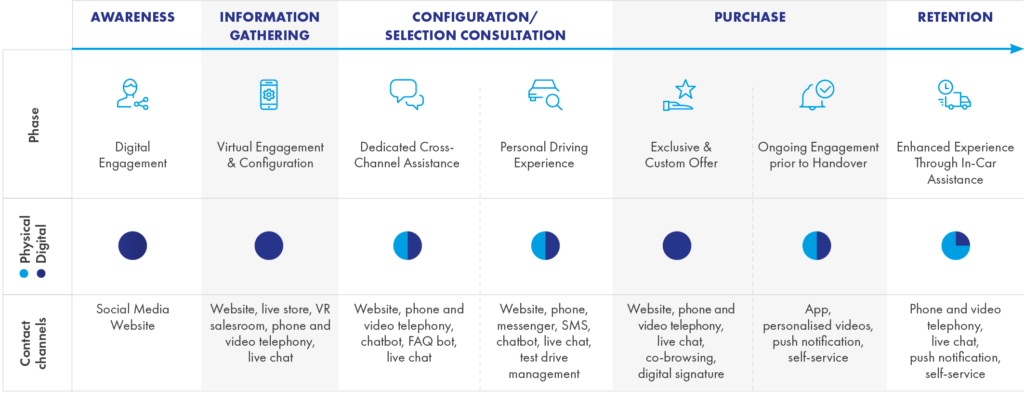
* + - Highly accurate Image Predictive model
    - better user responsive web application
    - Cloud database for storing the information’s

# PROJECT DESIGN

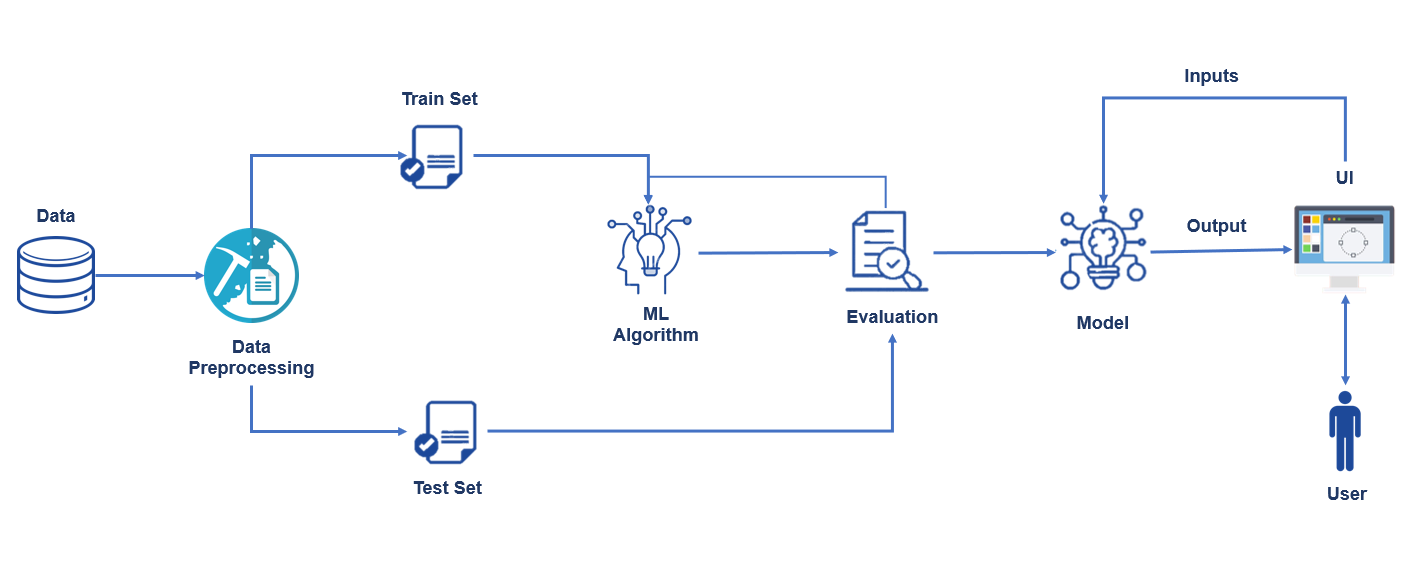
## Data Flow Diagrams

****

* 1. **Customer Journey Map**

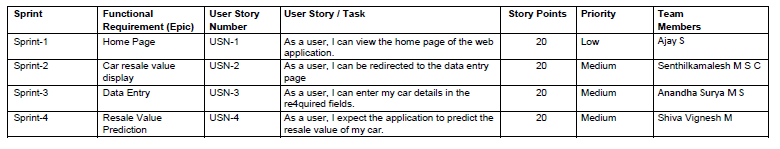


## Solution & Technical Architecture

****

# PROJECT PLANNING

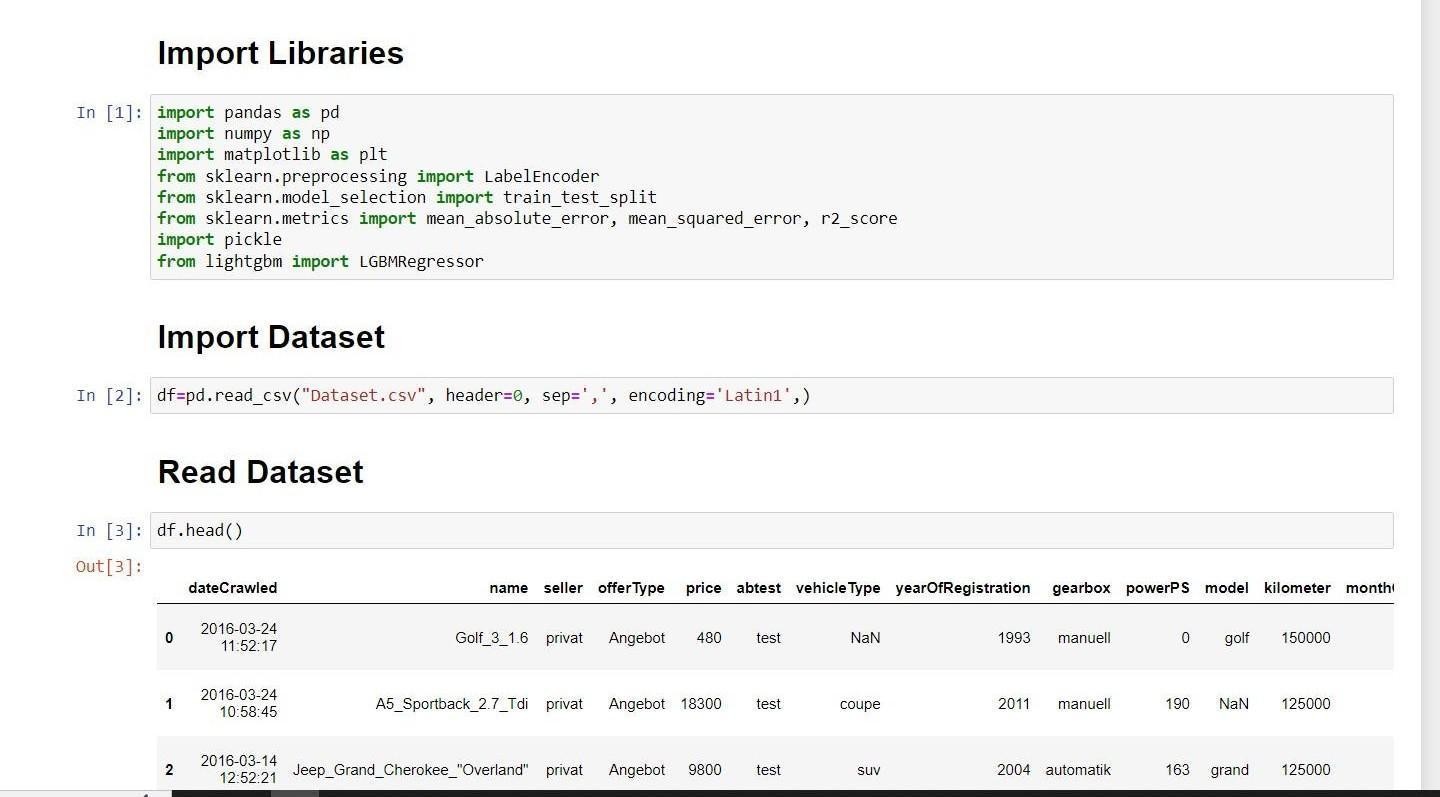
## Sprint Planning and Estimation

****

* 1. **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

# CODING & SOLUTION

## Feature 1



* 1. **Feature 2**

from flask import Flask, render\_template, Response, request import pandas as pd

import numpy as np import pickle

from sklearn.preprocessing import LabelEncoder app = Flask( name )

def load\_model(file='resale\_model.sav'): return pickle.load(open(file, 'rb'))

@app.route('/') def index():

return render\_template('homeview.html')

@app.route('/predict') def predict\_page():

return render\_template('predictionview.html')

@app.route('/y\_predict', methods=['GET','POST']) def predict():

regyear = int(request.args.get('regyear')) powerps = float(request.args.get('powerps')) kms = float(request.args.get('kms')) regmonth = int(request.args.get('regmonth')) gearbox = request.args.get('geartype') damage = request.args.get('damage')

model = request.args.get('model') brand = request.args.get('brand') fueltype = request.args.get('fuelType')

vehicletype = request.args.get('vehicletype')

new\_row = {'yearOfReg':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','yearOfReg','gearbox', 'powerPS','model','kilometer','monthOfRegistration','fuelType', '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'), allow\_pickle=True) transform = mapper[i].fit\_transform(new\_df[i]) new\_df.loc[:,i+'\_labels'] = pd.Series(transform, index=new\_df.index)

labeled = new\_df[['yearOfReg','powerPS','kilometer','monthOfRegistration'] + [x+'\_labels' for x in labels]]

X = labeled.values.tolist() print('\n\n', X)

predict = reg\_model.predict(X) print("Final prediction :",predict)

return render\_template('predictedview.html',predicted\_value=predict)

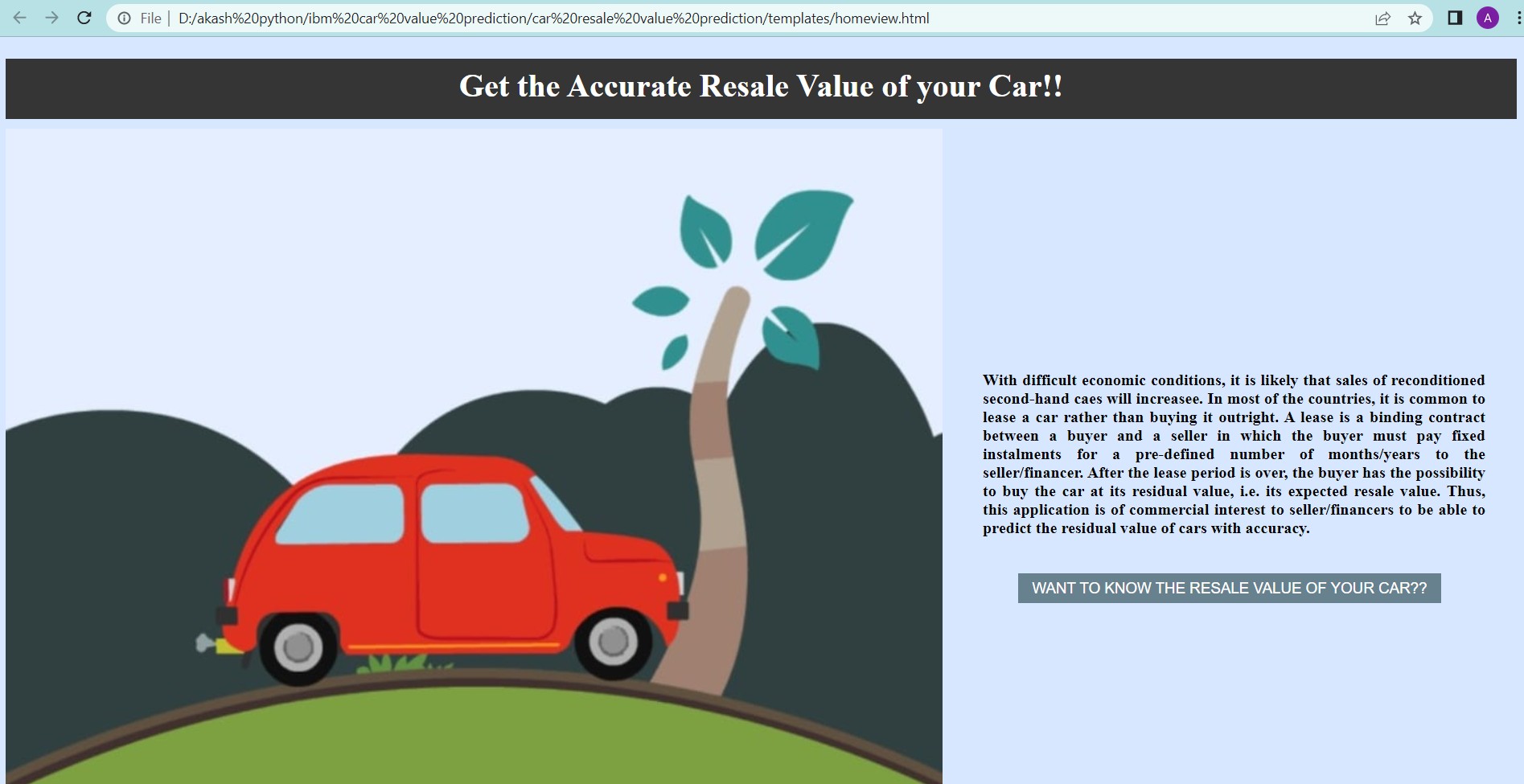
if name ==' main ': reg\_model = load\_model() app.run(debug=True)

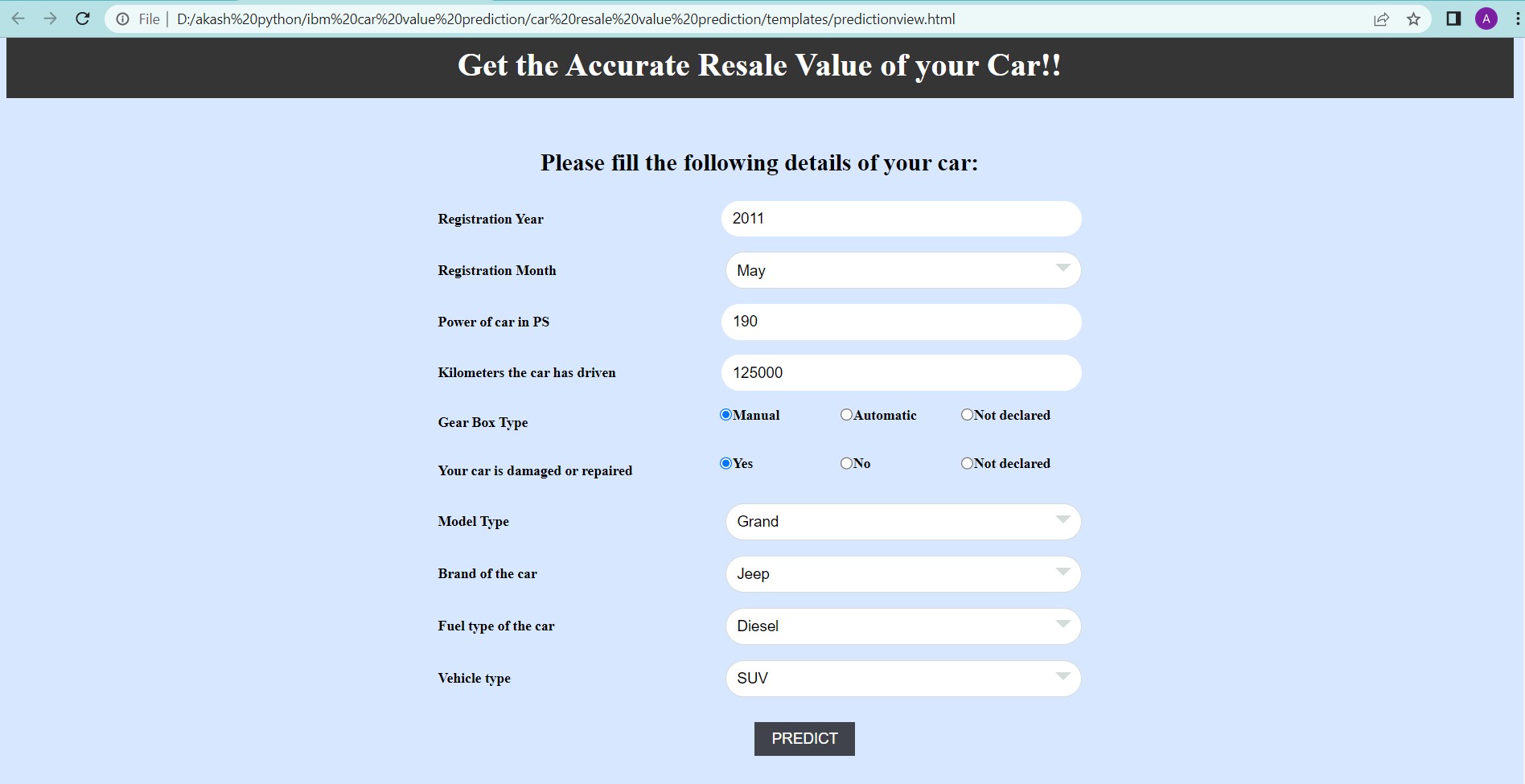
# TESTING

## Test Cases

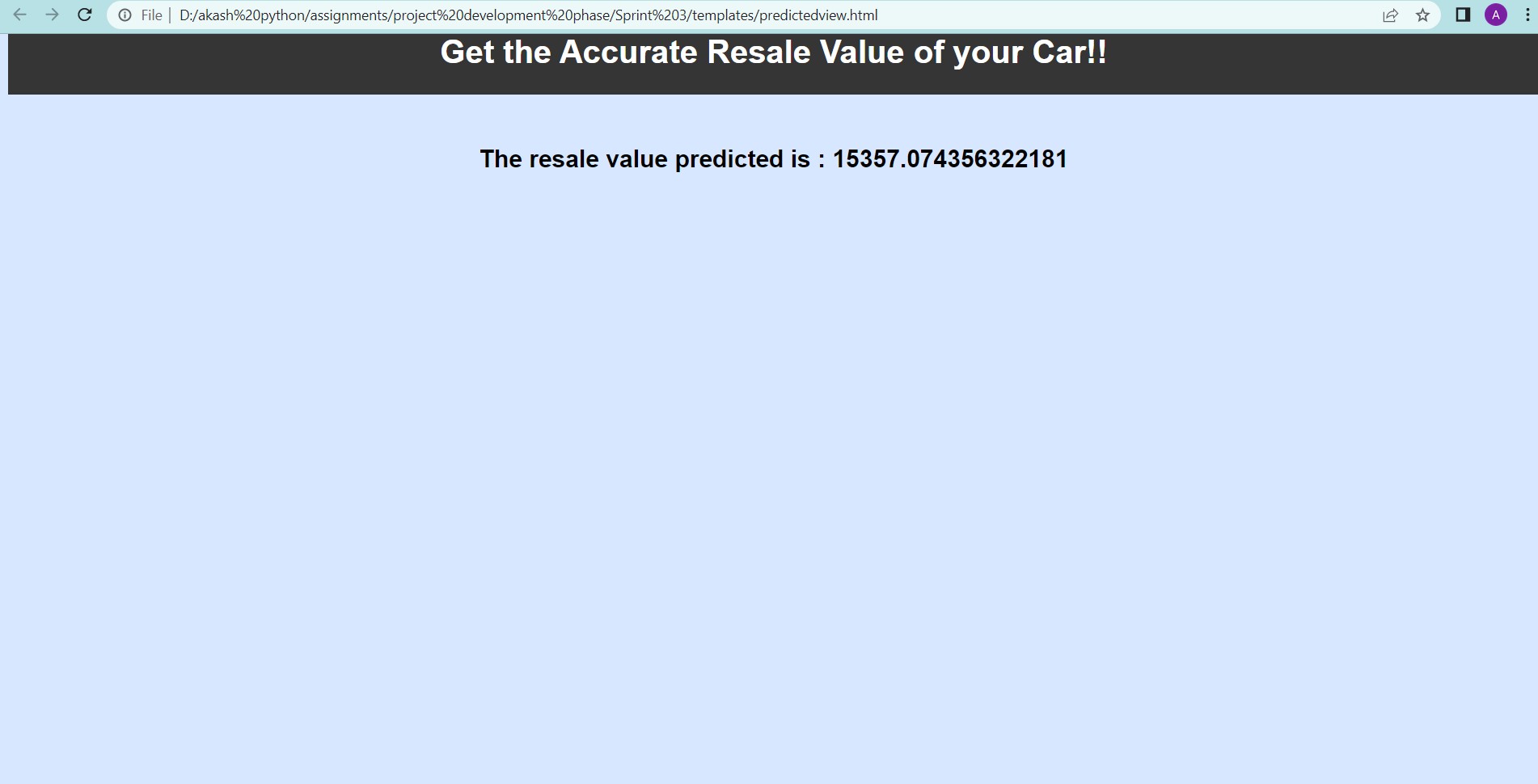
* + 1. User Login and Registration test
    2. Database Update test
    3. Prediction test

## User Acceptance Testing





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



# PERFORMANCE

## 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 are 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

# 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

# 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.

# 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-36137-1660293108**](https://github.com/IBM-EPBL/IBM-Project-36137-1660293108)

# APPENDIX

**App.py**

from flask import Flask, render\_template, Response, request import pandas as pd

import numpy as np import pickle

from sklearn.preprocessing import LabelEncoder app = Flask( name )

def load\_model(file='resale\_model.sav'): return pickle.load(open(file, 'rb'))

@app.route('/') def index():

return render\_template('homeview.html')

@app.route('/predict') def predict\_page():

return render\_template('predictionview.html')

@app.route('/y\_predict', methods=['GET','POST']) def predict():

regyear = int(request.args.get('regyear')) powerps = float(request.args.get('powerps')) kms = float(request.args.get('kms')) regmonth = int(request.args.get('regmonth')) gearbox = request.args.get('geartype') damage = request.args.get('damage')

model = request.args.get('model') brand = request.args.get('brand') fueltype = request.args.get('fuelType')

vehicletype = request.args.get('vehicletype')

new\_row = {'yearOfReg':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','yearOfReg','gearbox', 'powerPS','model','kilometer','monthOfRegistration','fuelType', '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'), allow\_pickle=True) transform = mapper[i].fit\_transform(new\_df[i]) new\_df.loc[:,i+'\_labels'] = pd.Series(transform, index=new\_df.index)

labeled = new\_df[['yearOfReg','powerPS','kilometer','monthOfRegistration'] + [x+'\_labels' for x in labels]]

X = labeled.values.tolist()

print('\n\n', X)

predict = reg\_model.predict(X) print("Final prediction :",predict)

return render\_template('predictedview.html',predicted\_value=predict)

if name ==' main ': reg\_model = load\_model() app.run(debug=True