PROJECT REPORT ON CAR RESALE VALUE PREDICTION
IBM

TEAM ID : PNT2022TMID32608

GIRI KARTHICK GR
JUFFIN HASSAN H
ASHWIN R
HARESH KRISHNARAJ R
ABDUL MAJEED AHMAR PEER IBRAHIM A

CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Project Overview :

In today's fast-paced world it's quite difficult to be patient. Industries are growing rapidly in such a manner that everyone must satisfy their needs as soon as possible. Having a car is a means of convenience of being able to travel anywhere you want . But due to the difficult economic conditions , people tend to look for used cars as a choice. In some countries, it is common to lease a car rather than buying it outright whereas in some countries a used car in fact makes more sense for first time buyers as it a value for money, upgrading from two-wheelers or public transportation, or for that matter, someone looking to buy a second set of wheels in the family.

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

Every car's value drops right from the moment it is bought and its depreciation continues with each passing year. In fact, in the first year itself, the value of a car decreases by 20 percent of its initial value. The make and model of a car, total kilometers driven, overall condition of the vehicle and various other factors further affect the car's resale value .

The increased prices of new cars and the financial incapability/instability of the customers leads into buying used cars. Therefore, in order to check the worthiness of the car as well satisfying vendor and customers we need a Car Price Prediction system which effectively determines the worthiness of the car using a variety of features.

CHAPTER 2
LITERATURE SURVEY

2.1 **Existing problem:**

The main challenge faced in any prediction is accuracy. Based on the distinct , important features the model should predict the more accurate data .Challenges like incompleteness in data, underfitting, overfitting ,extreme complexity ,Algorithmic imperfections leads to problems or act as an hindrance to predictive analysis.

2.2 References :

1. **Car Price Prediction Using Machine Learning - June 2021**

   Authors : Ketan Agrahari, Ayush Chaubey, Mamoor Khan, Manas Srivastava
   Technology and algorithms used:
   Machine Learning, Linear Regression, Lasso Regression, Correlation.
   The aim of this research paper is to predict the car price as per the data set (previous consumer data like engine capacity, distance travelled, year of manufacture, etc.). The result of these algorithms will be analysed and based on the efficiency and accuracy of these algorithms, the best one of them can be used for the said purpose.

2. **Used Cars Price Prediction using Supervised Learning Techniques – Dec 2019**

   Authors : Pattabiraman Venkatasubbu, Mukkesh Ganesh
   Technology and algorithms used:
   ANOVA, Lasso Regression, Regression Tree, Tukey's Test
   This analysis/paper focuses on a statistical model which will be able to predict the price of a used car, based on previous consumer data and a given set of features The emergence of online portals such as CarDheko, Quikr, Carwale, Cars24, and many others has facilitated the need for both the customer and the seller to be better informed about the trends and patterns that determine the value of the used car.

3. **Used car price prediction - Apr 2021**

   Authors : Praful Rane, Deep Pandya, Dhawal Kotak
   Technology and algorithms used:
   Machine Learning, Linear Regression, Ridge Regression, Lasso Regression, Decision Tree Regressor
   Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and it's value in the present day scenario. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at. To overcome this problem, a model was developed which is highly effective.

4. **How much is your car worth ? - Aug 2017**

   Author : Shonda Kuiper
   Technology and algorithms used:
   Machine Learning,Multiple Regression,Dummy Variables, Heteroskedasticity, Data Transformation,Residuals.

This paper discusses the development of a multivariate regression model to predict the retail price of 2005 General Motor (GM) cars. Statistical textbooks typically offer many small data sets chosen to illustrate a variety of issues and techniques that a user of regression should know. Although small data sets can offer the advantage of sharp focus on particular issues, their narrow focus carries disadvantages as well. Working with a large, richly structured data set can give students a kind of experience not possible with a succession of smaller data sets.

5. **How much is my car worth?**

A methodology for predicting used cars prices using Random Forest – Nov 2017

Authors : Nabarun Pal, Priya Arora, Dhanasekar Sundararaman, Puneet Kohli, Sai Sumanth Palakurthy

Technology used : Random Forests, Regression, Decision Trees

In this paper, a supervised learning method namely Random Forest is used to predict the prices of used cars. The model has been chosen after careful exploratory data analysis to determine the impact of each feature on price. A Random Forest with 500 Decision Trees were created to train the data. From experimental results, the training accuracy was found out to be 95.82%, and the testing accuracy was 83.63%. The model can predict the price of cars accurately by choosing the most correlated features.

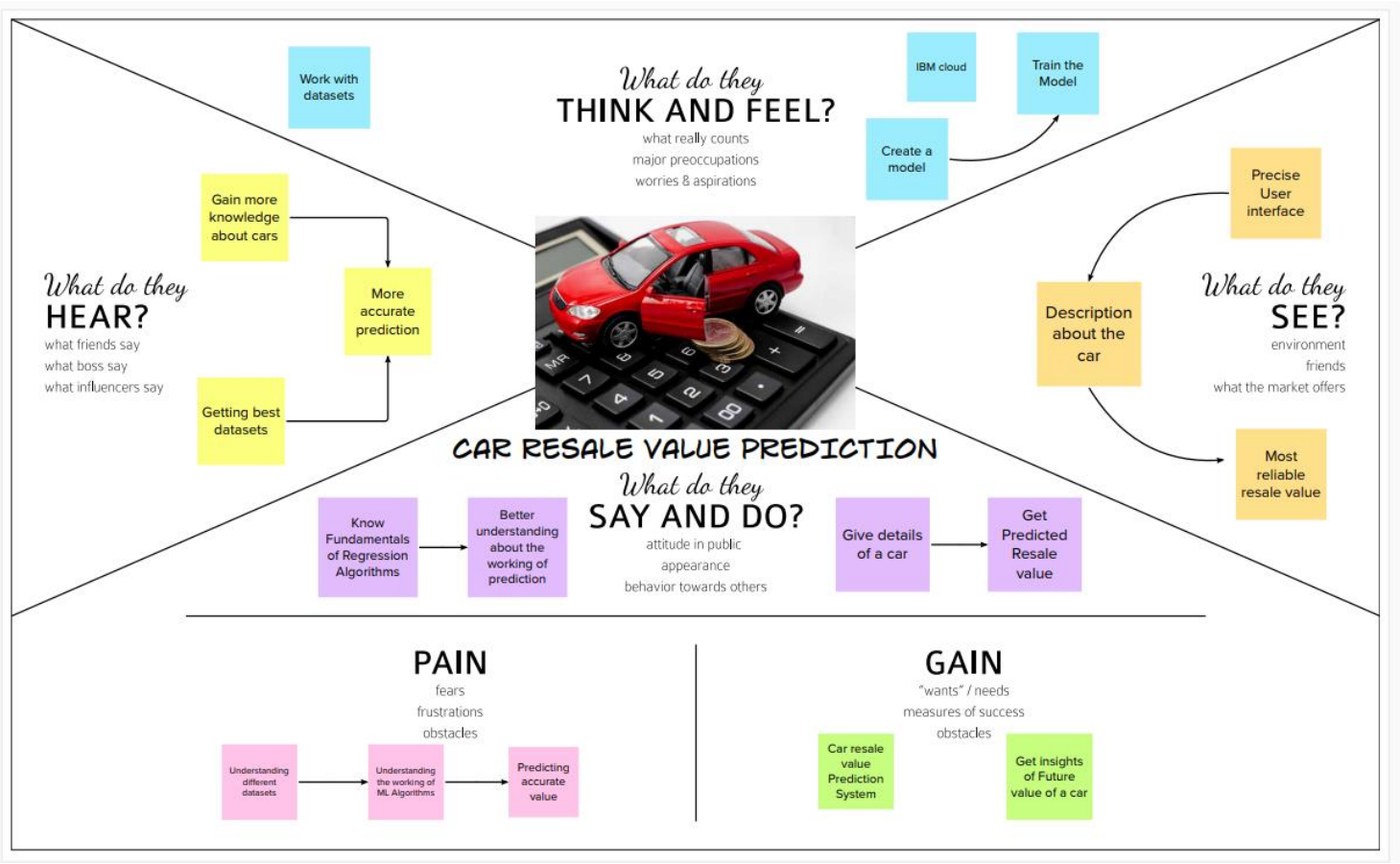
2.3 **Problem Statement Definition**:

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.

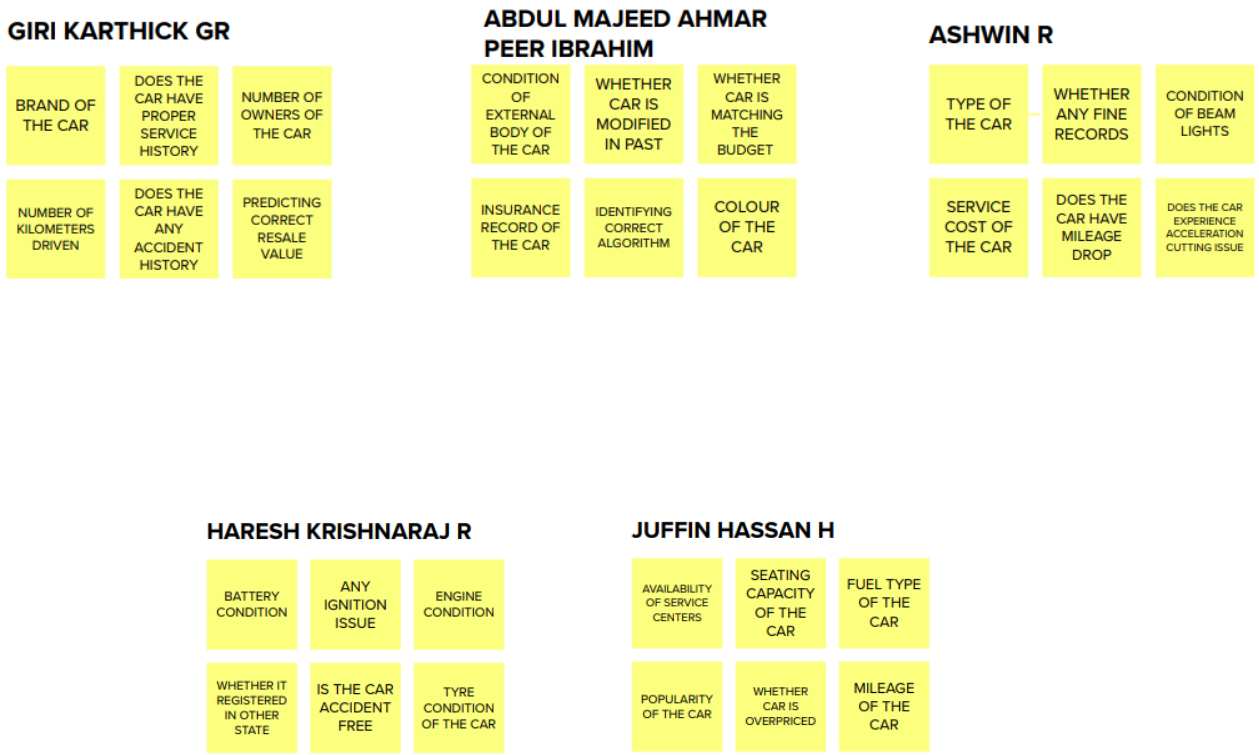
In order to predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale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

CHAPTER 3
IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



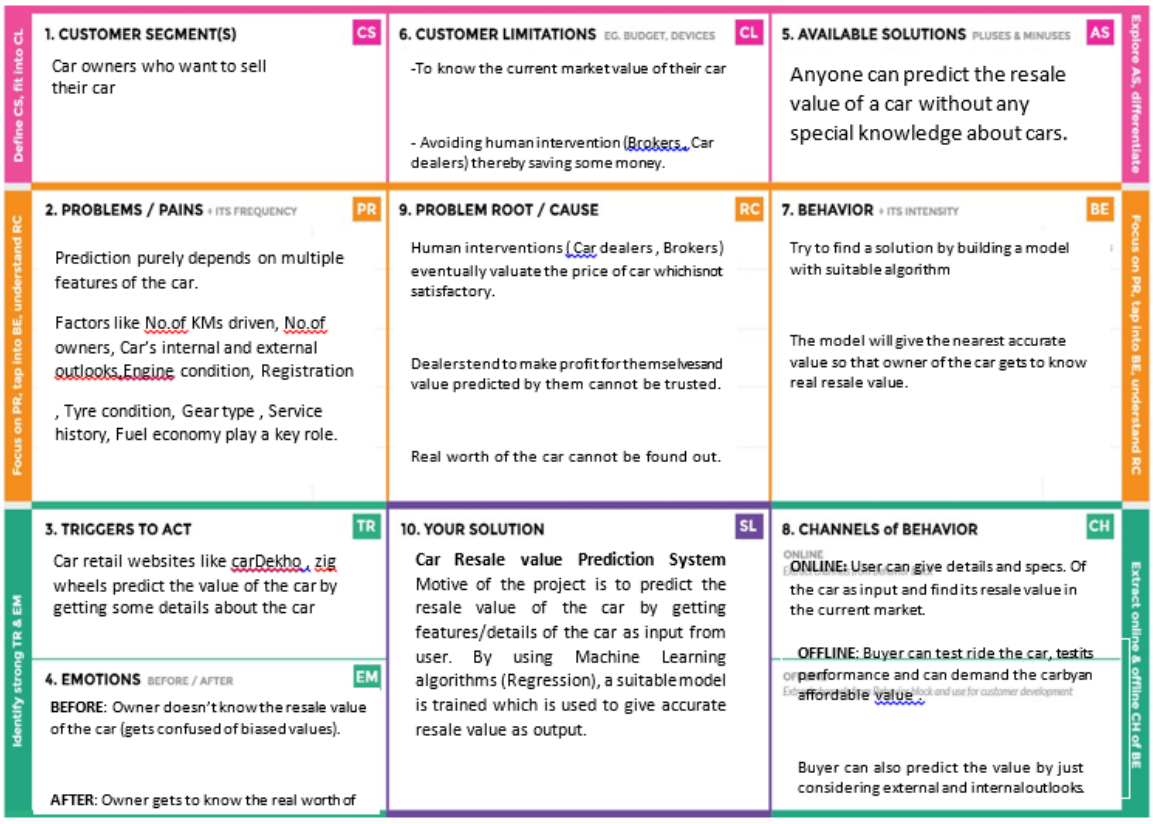
3.2 Ideation & Brainstorming



3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To develop a web application to predict the resale value of a car based on its features
2.	Idea / Solution description	To train the system with the dataset and train them using a regression model
3.	Novelty / Uniqueness	Predict more accurate value based on the input given by the user.
4.	Social Impact / Customer Satisfaction	Customer can get a clear recommendation of the resale value of the car and based on the evaluation , customer can take decisions on selling the car .
5.	Business Model (Revenue Model)	Free and can be accessed by anyone. Prediction purely depends on features of the car.
6.	Scalability of the Solution	More the specific details, more the accuracy. An accurate resale value is predicted by the trained model.

3.4 Problem Solution fit



CHAPTER 4
REQUIREMENT ANALYSIS

4.1 Functional requirements

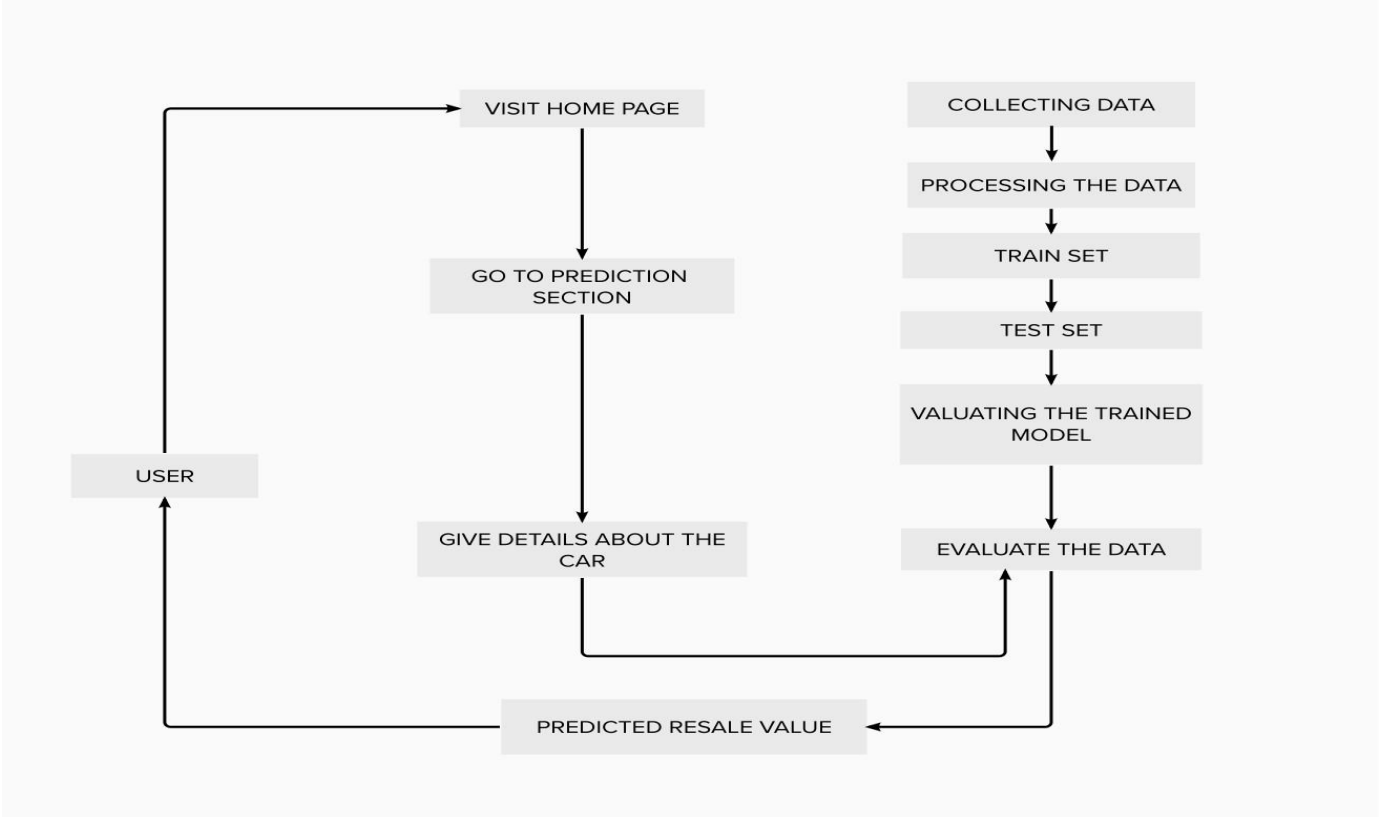
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Profile	Personal / General details of the user.
FR-4	User ‘s vehicle information	Features of the car such as Model , brand , Registration number, manufacturing year , fuel type , engine type

4.2 Non-Functional requirements

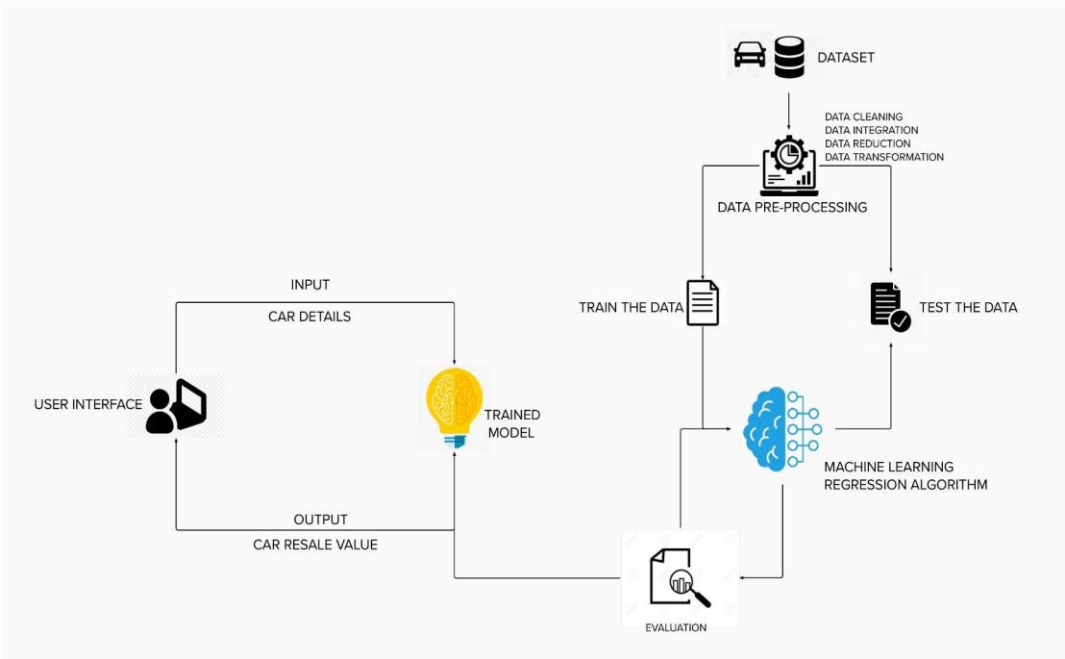
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	A Simple , User friendly user interface which can give solution to the user .(i.e resale value)
NFR-2	**Security**	It is secured and encrypted application with no spam redirects or hyperlinks.
NFR-3	**Reliability**	Reliable and consistent prediction can be done .
NFR-4	**Performance**	Performance does not depend on amount of end users. It is scalable , consistent with no delay
NFR-5	**Availability**	Uninterrupted services must be available other than any server issues .
NFR-6	**Scalability**	It can handle any amount of data irrespective of number of users. It can perform computations in a fast & cost effective manner.

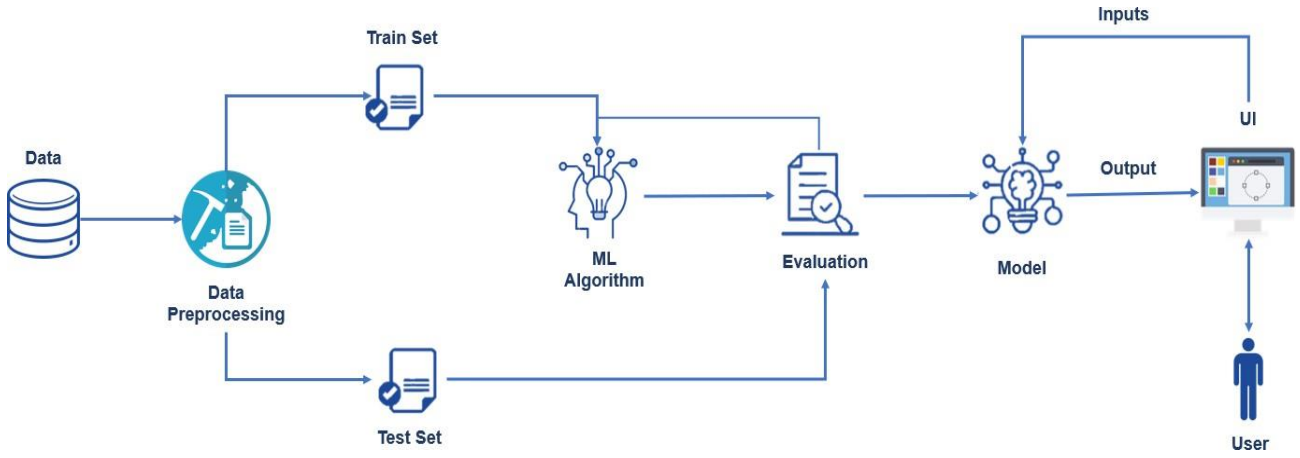
CHAPTER 5
PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture





5.3 User Stories

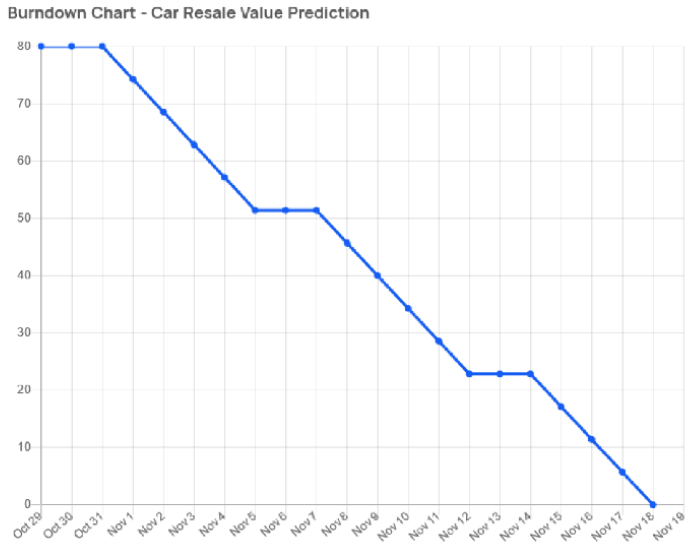
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (User)	Visit the website	USN-1	As a user, I can visit the home page of the website and get to know about the website.	I can access the website by correct url .	High	Sprint-1
	Predict the value of my old car	USN-2	As a user, I can predict the resale value of the car by giving its details/ features .	I can enter car details	High	Sprint-1
	Look for a price of an used car	USN-3	As a user and also a buyer or seeker , I can also get to know the price of the used car .	I can choose most likely car.	Low	Sprint-1
Customer Care Executive	Application	USN-4	As an executive I can clarify doubts of users by using chat bots etc.	I can provide support to the users.	High	
Administrator	Application	USN-5	As an administrator I can upgrade / update /modify the application with additional features.		High	

CHAPTER 6
PROJECT PLANNING & SCHEDULING
6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Home page	USN-1	As a user, I can view the homepage of the web application	20	Low	Giri karthick GR
Sprint-2	Entry of Car Details & Specifications	USN-2	As a user, I can input/entry the car details in the Data Entry Page.	10	Medium	Haresh Krishnaraj R
Sprint-2	Modification of Car Details	USN-3	As a user, I can modify the car details and specifications if they are ignored/Missing	10	High	Abdul majeed ahmar peer Ibrahim. A
Sprint-3	Display Car resale value	USN-4	As a user, I can view the current market resale value of the car.	20	High	Ashwin R
Sprint-4	Car Resale Value Prediction	USN-5	As a user, I expect the most accurate genuine resale value of car with consideration of all external factors.	20	High	Juffin Hassan H

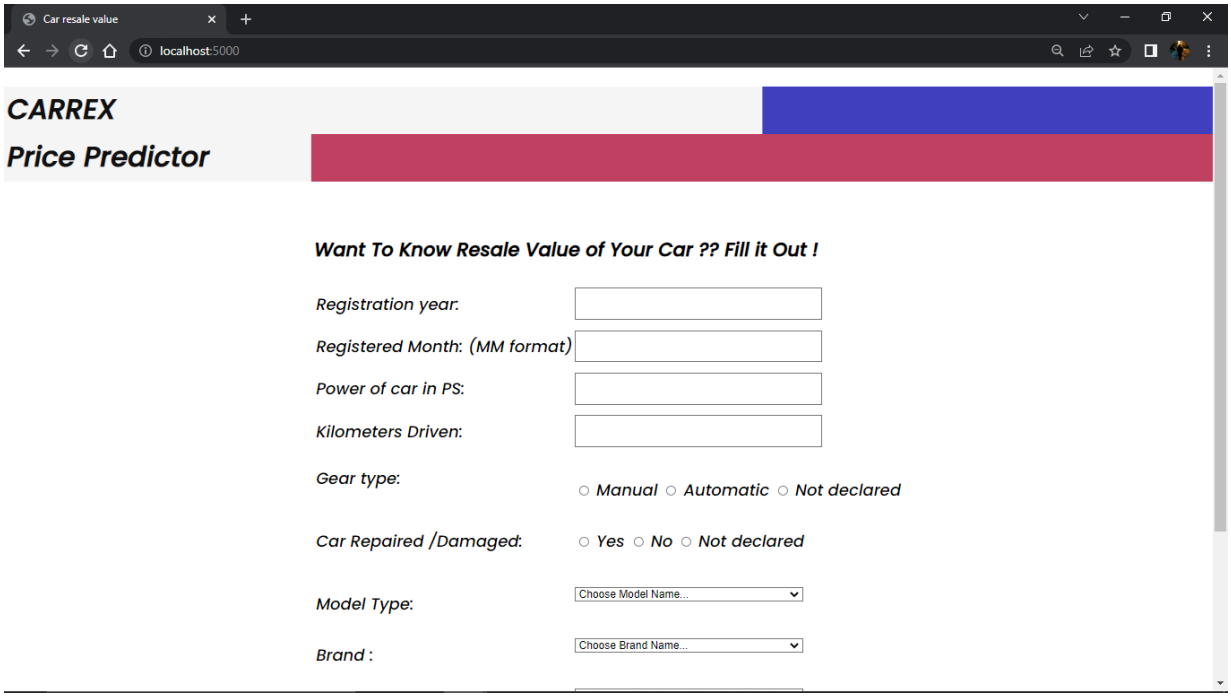
6.2 Reports from JIRA

Burndown Chart:

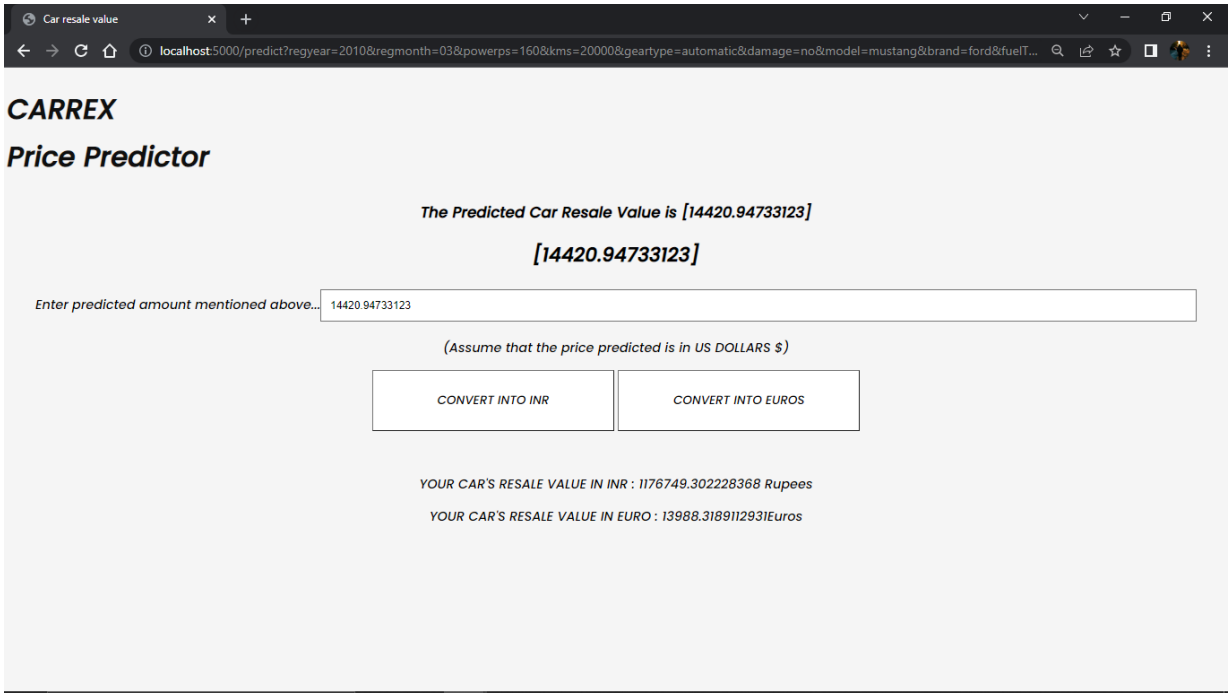


CHAPTER 7
CODING & SOLUTIONING

Feature 1 : CAR DETAILS / DATA ENTRY PAGE



Feature 2 : PREDICTION PAGE



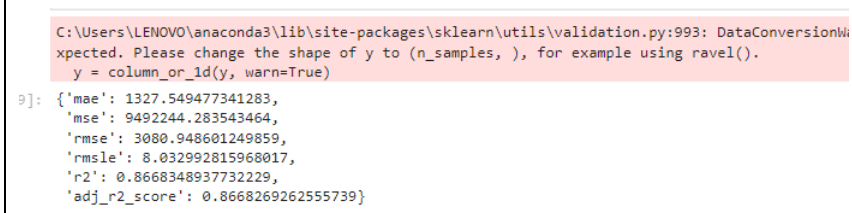
Flask Application- > app.py

```
import pandas as pd
import numpy as np
from flask import Flask, render_template, Response, request
import pickle
from sklearn.preprocessing import LabelEncoder
import requests

app = Flask(__name__)
def load_model(file='resale_model.sav'):
    return pickle.load(open(file, 'rb'))

@app.route('/')
def index():
    return render_template('value.html')

@app.route('/predict_page')
def predict_page():
    return render_template('value.html')

@app.route('/predict', methods=['GET', 'POST'])
def predict():
    reg_year = int(request.args.get('regyear'))
    powerps = float(request.args.get('powerps'))
    kms = float(request.args.get('kms'))
    reg_month = 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')
    fuel_type = request.args.get('fuelType')
    veh_type = request.args.get('vehicletype')
    new_row = {'yearOfReg': reg_year, 'powerPS': powerps, 'kilometer': kms,
            'monthOfRegistration': reg_month, 'gearbox': gearbox,
```

```
            'notRepairedDamage': damage,
            'model': model, 'brand': brand, 'fuelType': fuel_type,
            'vehicletype': veh_type}
    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)
    payload_scoring = {"input_data": [{"fields": [['yearOfReg', 'powerPS', 'kilometer',
'monthOfRegistration', 'gearbox_labels',
                                                   'notRepairedDamage_labels', 'model_labels',
'brand_labels', 'fuelType_labels', 'vehicletype_labels']], "values": X}]}
    print("Final prediction :", predict)
    return render_template('predict.html', predict=predict)
if __name__ == '__main__':
    reg_model = load_model()
    app.run(host='localhost', debug=True, threaded=False)
```

CHAPTER 8
TESTING

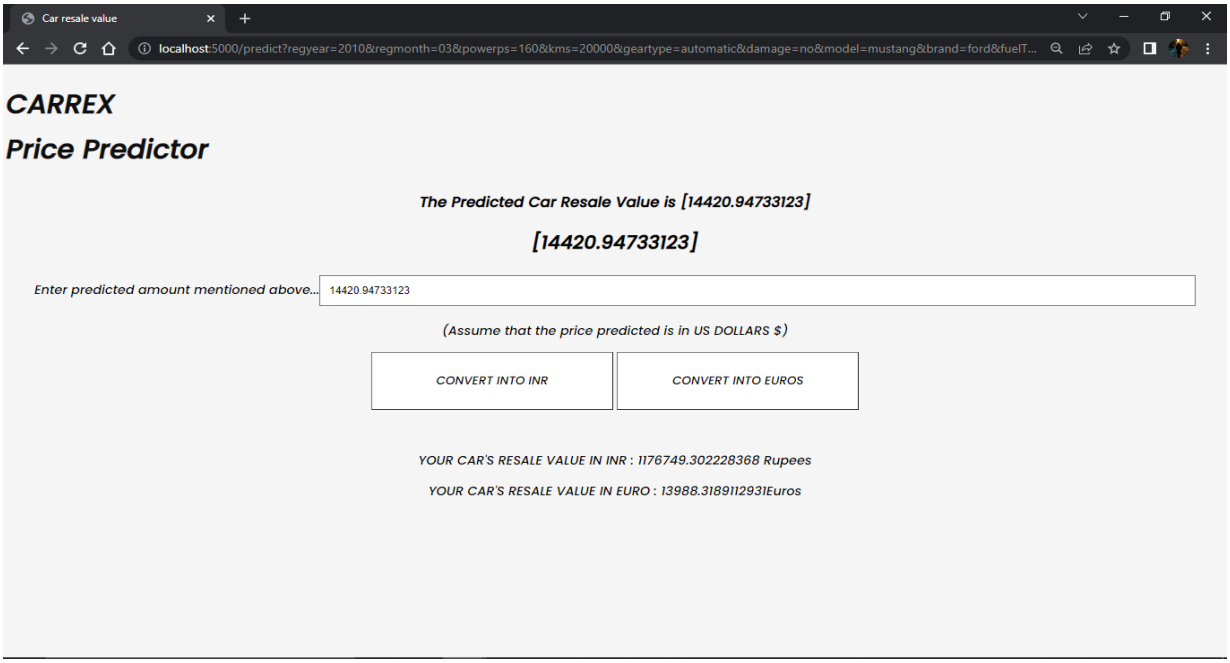
8.1 **Performance Testing:.**

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Regression Model:** MAE - , MSE - , RMSE - , R2 score - **Classification Model:** Confusion Matrix - , Accuray Score- & Classification Report -	
2.	Tune the Model	Hyperparameter Tuning - Validation Method -	

User Acceptance Testing:

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	5	2	3	20
Duplicate	1	2	3	10	4
External	210	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	78

CHAPTER 9
RESULTS



Performance Metrics :There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are:

- Mean Squared Error (MSE).
- Root Mean Squared Error (RMSE).
- Mean Absolute Error (MAE)

```
    'model_labels', 'brand_labels', 'fuelType_labels',
    'vehicleType_labels'],
   dtype='object')
```

Different Metrics Evaluation

In [6]:
```
def find_scores(Y_actual, Y_pred, X_train):
    scores = dict()
    mae = mean_absolute_error(Y_actual, Y_pred)
    mse = mean_squared_error(Y_actual, Y_pred)
    rmse = np.sqrt(mse)
    rmsle = np.log(rmse)
    r2 = r2_score(Y_actual, Y_pred)
    n, k = X_train.shape
    adj_r2_score = 1 - ((1-r2)*(n-1)/(n-k-1))

    scores['mae']=mae
    scores['mse']=mse
    scores['rmse']=rmse
    scores['rmsle']=rmsle
    scores['r2']=r2
    scores['adj_r2_score']=adj_r2_score

    return scores
```

Train Test Split

In [7]:
```
X = labeled.iloc[:,1:].values
Y = labeled.iloc[:,0].values.reshape(-1,1)
```

In [8]:
```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.4, random_state=42)
```

CHAPTER 10

ADVANTAGES & DISADVANTAGES OF PRICE PREDICTION SYSTEM

10.1 ADVANTAGES

10.1.1 Brand: Some brands have prolonged value due to their popularity, while others lose their value significantly due to increasing competition. Thus, thorough research can get you close to 60 percent of a car's price in three years.

Performance: A car offering better distance per liter of fuel is your best bet as it will not only save you money but is also an indicator of a car's good performance.

10.1.2 Decent condition of used cars :

Gone are the days when used cars meant shabby, worn-out exteriors and interiors and scratches all over. The users in today's age do not have to sacrifice reliability and overall condition to strike a good deal on a used car. There are plenty of options in the used cars segment and you can select a car that is scratch-free and in good mechanical shape. You can easily finalize one that is "looks like new". Moreover, when you purchase from big dealers, certified pre-owned vehicles with a manufacturer's warranty often meet higher detailing, appearance and mechanical standards.

10.1.3 Affordableprices :

Old cars are always much cheaper than new cars but there is always a problem of inheriting the problems faced by the previous owner. This can be easily negated by checking whether which of the car is certified or not. If you have always dreamt of owning a luxury car but are not able to do so because of budget problems, you can easily enjoy the pleasure of driving a luxury car by buying a second hand model.

10.1.4 Low Depreciation Rates

A newly acquired car depreciates at a higher rate than an old car and that is where you are set to gain. Cars lose some value with each passing month and mile. But the highest loss in value occurs in the first year and it is close to 40%. When buying an old car, you do not have to face any such huge depreciation. Also, there is less mental depreciation since you do not have to worry about the rock chip in the paint or about the parking-lot ding.

10.1.5 Low Insurance Rates

Just similar to financing, insurance rates are also affected by the age of a car. But in case of used cars, insurance rates tend to be less expensive. Those who do a little bit of pre-purchase research get saved from the insurance sticker shock, irrespective of which vehicle they choose.

10.1.6 Warranty:

When you purchase a used car from a company owned used car outlet, you also get a warranty on the vehicle. However, this warranty comes with a certain limit and it covers certain kms that you travel in a stipulated time.

10.1.7 Reduced rates for car insurance premiums

The motor insurance premiums are calculated on the basis of the market value of the vehicle. This is why the insurance companies charge higher premiums for insurance policies of brand-new car owing to their high market value.

On the flip side, the value of a pre-owned car would have already undergone a significant depreciation over the years and hence, you can get it insured at a much lower price by quoting a nominal IDV or Insured Declared Value for the vehicle.

10.1.7 Sales Tax

Most ads for cars usually gloss over the tax issue. Many state laws levy taxes on purchase of new cars but no tax is levied on used cars. This way the buyers of old cars can save on a lot of money

DISADVANTAGES :

1. Higher Interest Rates

   you need to apply for a loan of a lesser amount just because the price of the car is low. However, banks and loan providers usually charge a higher rate of interest on a used car loan than they do on a new car loan. Ther need to earn more anyhow so here's the catch. You might save on the principle amount but can't save on the interest rates.

2.Lack of choice or Additional Benefits

   New car manufacturers have to get rid of the stock at some particular time so they offer you huge discounts. The case with the used car market is a little different. There is no special month, they don't have to worry about the year ends too. Although the prices are obviously lower in the used car markets, you don't get that much "discounts".

CHAPTER 11

CONCLUSION

Pricing a car is crucial for buying and selling a car. Considering the fact that underpricing your car can lead to you losing money on selling it. Also, you'll receive a lower claim payout under your insurance policy for a total loss event like car accident/ theft.Car valuation/prediction helps the sellers get an idea of what their car is worth and how much they should expect to get for their car. When it comes to the buyers, they too get an idea of what's the maximum amount they should be paid while buying a particular car. This way the tool proves to be beneficial for both parties.

12. FUTURE SCOPE:

- Car prediction system can be further improvised by increasing more and valid data.
- Car prices across various countries or resale value can be checked.

CHAPTER 13 APPENDIX

Source Code

GitHub Link : https://github.com/IBM-EPBL/IBM-Project-16665-1659619861

Demo link: https://www.youtube.com/watch?v=wbx1lwyao5I