Project Report

IBM Naliya Thiran

Car Resale Value Prediction

Dr. Mahalingam College of Engineering and Technology

Team ID	PNT2022TMID37777
Domain	Applied Machine Learning
Team Leader	Silambarasan G
Team Members	Krishnaraj E Prasanth G Vijay K

Contents

1 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 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.

1.1 Project Overview

A car price prediction has been a high-interest research area, as it requires noticeable effort and knowledge of the field expert. A considerable number of distinct attributes are examined for 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

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

1.2 Purpose

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

2. LITERATURE SURVEY

2.1 Existing problem

The problem is defined as the optimized 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 car 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 the power of 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.

S. no	Author, Title, Publication	Tech Stack	Advantage	Disadvantage
01	Praful Rane, Deep Pandya, Dhawal Kotak, 'USED CAR PRICE PREDICTION', International Research Journal of Engineering and Technology (IRJET) Published on April 4, 2021	Random Forest Regression, Hyperparameters, Random forest, Extra-tree Regressor	It is estimated that a car's value depreciated to almost 50%. The calculation for the first year is about 20%. That means if you buy a used car launched a year ago, you can save up to 20% on its original cost.	Even though we curated a list of available cars under 50,000 but that wasn't for you to buy and use for your daily commute. Usually, these cars are lemons. They look fine on the outside but can land you huge repair costs while you use them.

02	Aditya Nikhade, Rohan Borde, Car Price Prediction using Machine Learning, International Advanced Research Journal in Science, Engineering and Technology Published on April 4, 2022	Linear Regression, Ridge Regression, and Lasso Regression	But the insurance premium for a new car is comparatively higher than an old car. If you search for a good vehicle in the used car market, the previous owner has already paid hefty premiums when the car was new.	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. Their need to earn more anyhow so here's the catch.
03	Mukkesh Ganesh, Pattabirama Venkatasubbu, 'Used Cars Price Prediction using Supervised Learning Techniques', International Journal of Engineering and Advanced Technology Published in December 2019	ANOVA, Lasso Regression, Regression Tree, Tukey's Test	If you buy a used car from a brand authorized dealership, you get a warranty on the repair. Read this carefully. In my case, I bought a Maruti Suzuki car from a True Value dealership and the car was still under its original warranty from the factory.	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-end too. Although the prices are obviously lower in the used car markets, you don't get that many "discounts".
04	Sameerchand Pudaruth 'Predicting the Price of Used Cars using Machine Learning Techniques' International Journal of Information & Computation Technology Published In 2014	Multiple linear regression analysis, k-nearest neighbors, naive Bayes and decision trees	This is the major difference between a new car and a used car. For the same cost as a new car, you can buy a used car from a couple of segments above. Imagine getting a D-segment car for the price of a top-end premium hatchback.	While you will be saving on lesser down payment and insurance costs, a used car comes with years of wear and tear in its parts and thousands of kilometers in its odo.

05	Abhishek Jha, Dr. Ramveer Singh Manish, Imran Saifi, Shipra Srivastava, 'Used car price prediction' International Journal of Advance Research, Ideas and Innovations in Technology(IJARIIT) Published in July 19, 2021	Python, Flask, Pandas, Numpy, Seaborn, Sklearn, Matplotlib, Random Forest Regressor	While it's true that it's harder to get financing as easily as a new car, you get to pay a lot lesser in the case of used cars. The same goes for insurance too as you will be paying a lesser insurance bill.	When you buy a new car, it's made to order. You have the option of picking the color, the features, whether or not you want a sunroof, and more. When you buy a used car, you get what you pay for. If the car has a crappy radio, you'll have to deal with it or pay to have it replaced.
06	Ashutosh Datt Sharma, Vibhor Sharma, Sahil Mittal, Gautam Jain, Sudha Narang, 'Predictive Analysis of Used Car Prices Using Machine Learning, International Research Journal of Modernization in Engineering Technology and Science Published in June 6, 2021	Random Forest Regression (Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Plotly, Pickle, HTML, CSS, Flask, Jsonify)	That's because modifying it during the 3- to the 5 years since it was bought will void the warranty. If the 2nd hand vehicle you plan to buy is past that period, you're now free to "pimp" your ride.	Vehicles went from having CD players to auxiliary plugs to Bluetooth connections. When you purchase a used car, there's a chance you'll be getting older and less reliable technology.

07	K.Samruddhi, Dr. R.Ashok Kumar, 'Used Car Price Prediction using K-Nearest Neighbor Based Model' International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE) Published in September 2020	K Nearest Neighbor Regression Crossvalidation , K-Fold	When you decide to sell your car, the depreciation value is lesser on a used car than a new one, as most of the depreciation happens in the first two or three years.	While you will be saving on lesser down payment and insurance costs, a used car comes with years of wear and tear in its parts and thousands of kilometers in its odo. That equals more time and money for maintenance
08	Mr. Ram Prashath R, NIthish C N, Ajith Kumar J, 'Price Prediction of Used Cars Using Machine Learning' International Journal For Research in Applied Science and Engineering Technology Published in May 28,2022	Lasso Regression, Ridge Regression, Linear Regression	In the used car market, this strategy can benefit vendors, purchasers, and car manufacturers. It can then produce a reasonably accurate price estimate based on the data that users provide	Due to the huge requirement of used cars and the lack of experts who can determine the correct valuation, there is an utmost need of bridging this gap between sellers and buyers

09	R. Uma, J. Kamal, G. Sri Siva Thandavan, S. Raghul, 'Forecasting Vehicle Prices using Machine Learning Techniques', INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Published on June 15, 2022	Random Forest, XGBoost, LightGBM, CatBoost, Extra Trees	In my case, I bought a Maruti Suzuki car from a True Value dealership and the car was still under its original warranty from the factory. So I got dual benefits. will get True Value's additional warranty.	Find a car that has got a good engine and suspension but lacks features because it's a lower variant. On the other hand, you find a higher variant model that has not been maintained properly a bit challenging.
10	Enis Gegic, Becker Isakovic, Dino Keco Zerina Masetic, Jasmin Kevric, 'Car Price Prediction using Machine Learning Techniques', TEM Journal Published on February 16, 2019	Artificial Neural Network, Support Vector Machine, and Random Forest	The most obvious advantage among pros and cons of buying a used car is that you get the car at a cheaper price, therefore, you have to apply for a loan of a lesser amount and therefore, you are liable to pay lesser interest on that too	The manufacturer is not responsible for that. Provided, if the car comes with an already existing warranty straight from the manufacturer, you can get it transferred but you won't get any warranty/guarantee if you buy from an individual

2.3 Problem statement definition

3. IDEATION& PROPOSED SOLUTION

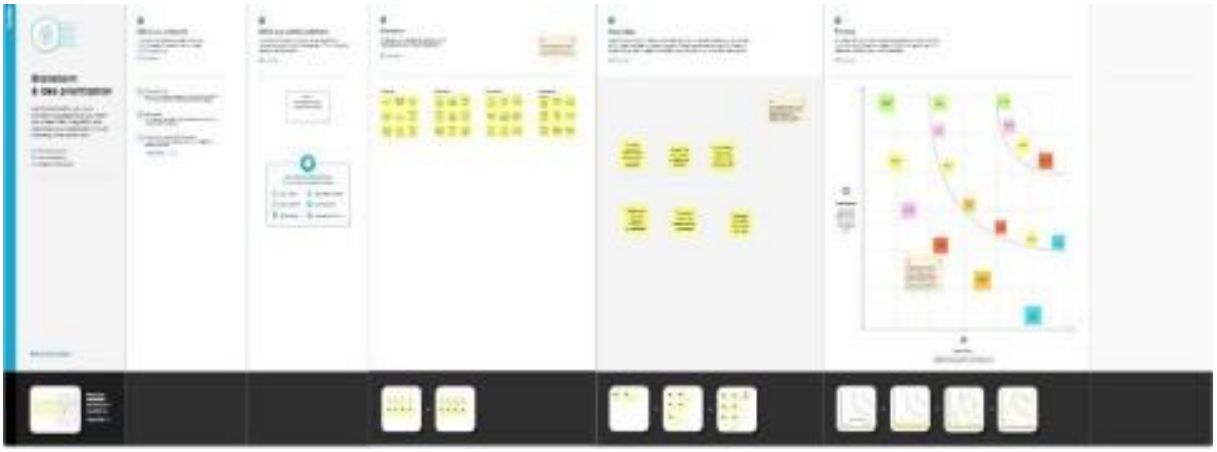
3.1 Empathy Map Canvas

3.2 Ideation & Brainstorming

- STEP 1
    - Prediction using Car image.
    - By using the exterior and interior images of the car.
    - The value will be predicted based on the appearance of the car.If there is any damage or n numbers scratches the car's resale value will be quite affected.
    - By using the 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.
    - Insurance, Company claims, etc.
    - a 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 the audio file.
- By using Convolutional Neural Networks methodology price can be predicted.
- CNNs for Machine Learning on sound data by spectrogram approach that just converts each song (or song segment) into a spectrogram: a two-dimensional matrix

- STEP 4
  - Economic Conditions.
  - Kilo-metres Covered.
  - Its mileage (the number of kilometers it has run) and its horsepower
  - Car prediction using XGBoost algorithm accurate results will be monitored. o XGBoost as a regression model gave the best MSLE and RMSE values.

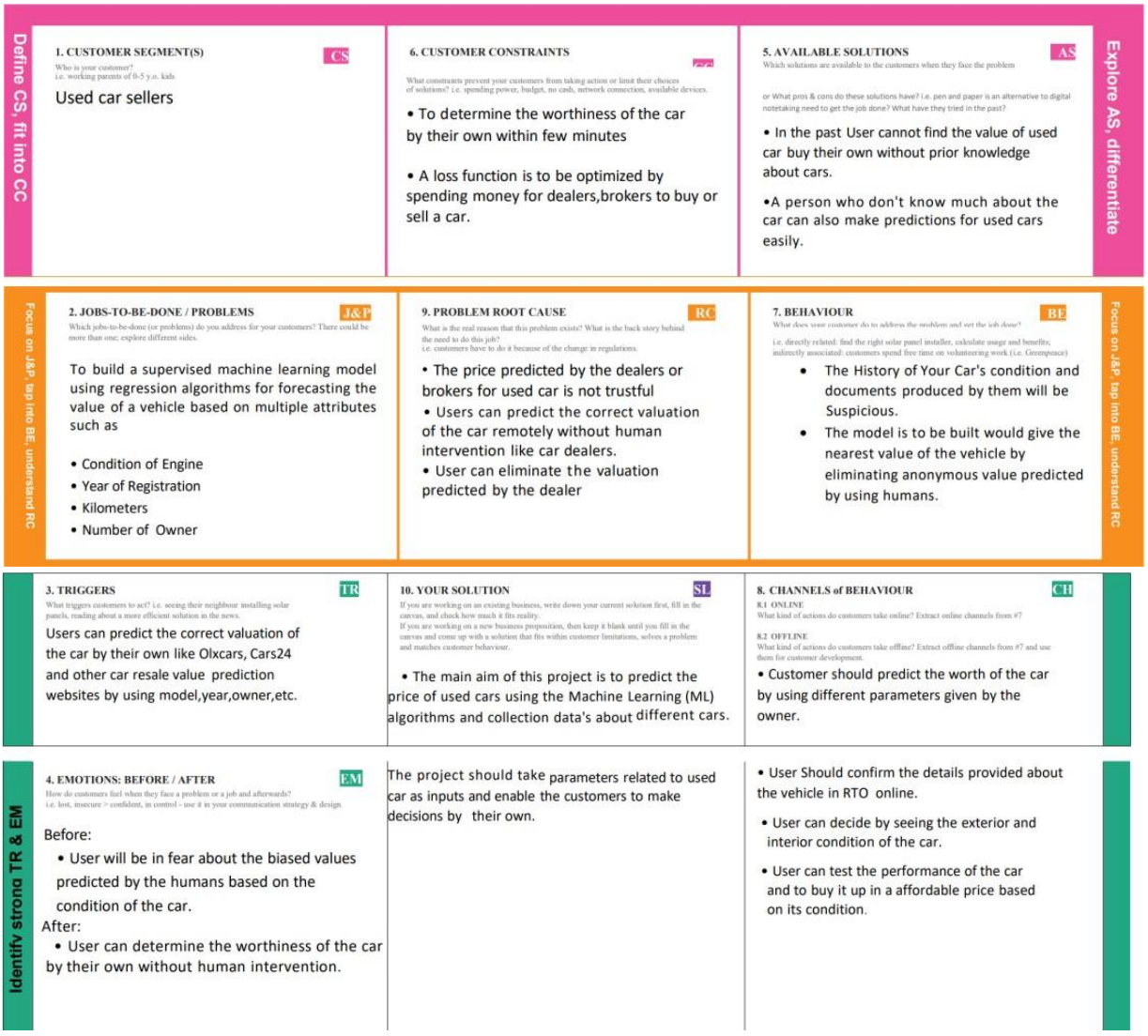


3.3 Proposed Solution

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 less amount of 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 a few seconds
6.	Scalability of the Solution	The solution given by the trained system is efficient and is a nearly accurate value of the vehicle.

3.4 Proposed Solution fit



4 REQUIREMENT ANALYSIS

4.1. Functional Requirement

The functional Requirements of this project involve a 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
- Pycharm
- PIP 2.7
- Jupyter Notebook
- Chrome

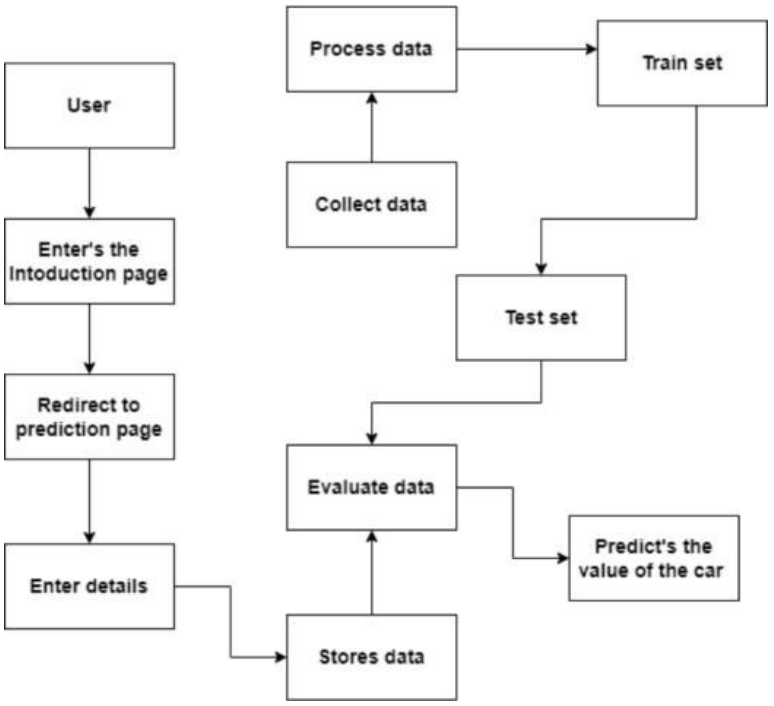
4.2. Non-Functional Requirement

The Non - Functional Requirements of this project are,

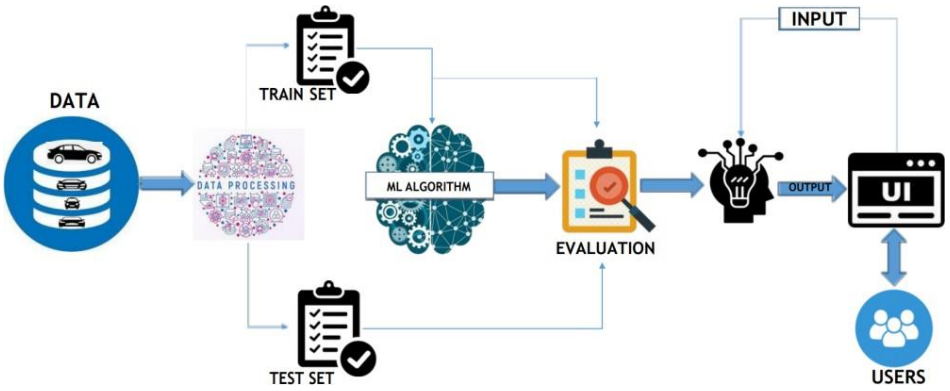
- Highly accurate Image Predictive model
- Better user-responsive web application
- Cloud database for storing the information

5. PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture



5.3 User Stories

Sprint	Functional Requirement (Epic)	User Story No.	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	Data Entry	USN-3	As a user, I can enter my car details in the required fields.	20	Medium
Sprint-4	Resale Value Prediction	USN-4	As a user, I expect the application to predict the resale value of my car.	20	Medium

6. PROJECT PLANNING

6.1 Sprint Planning and Estimation

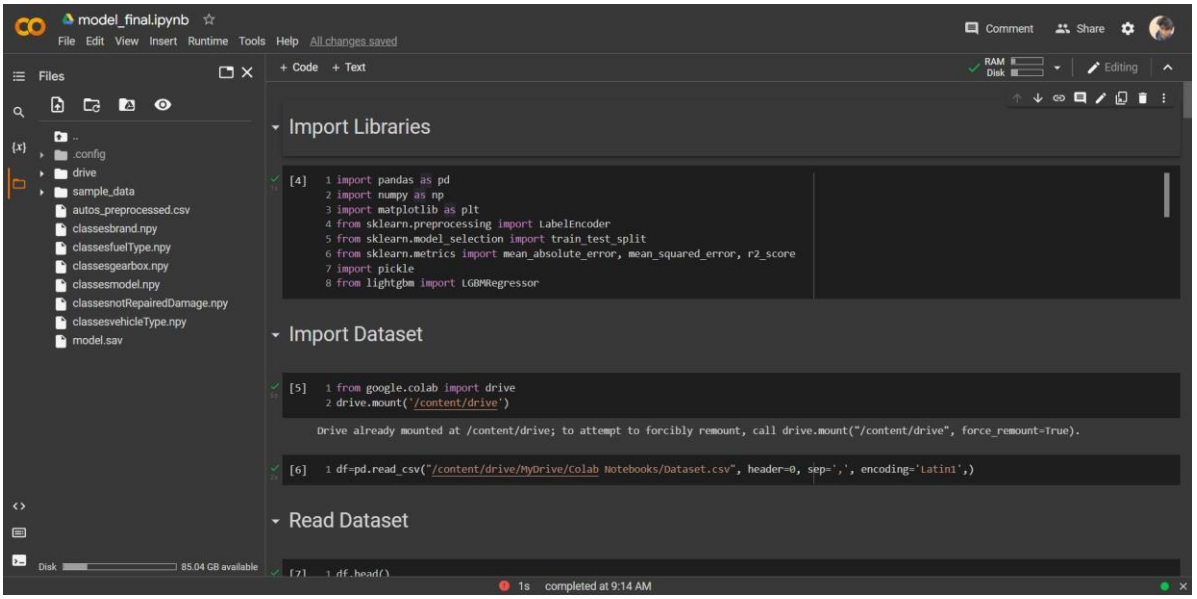
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

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

7.1 Feature 1



7.2 Feature 2

```
# Import Libraries

import pandas as pd

import numpy as np

from flask import Flask, render_template, Response,

request

import pickle

from sklearn.preprocessing import LabelEncoder
```

```
app = Flask(__name__)#initiate flask app


def load_model(file='model.sav'):#load the saved
model
    return pickle.load(open(file, 'rb'))


@app.route('/')
def index():#main page
    return render_template('car.html')


@app.route('/predict_page')
def predict_page():#predicting 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','ge
arbox',


'powerPS','model','kilometer','monthOfRegistration','f
uelType',
        'brand','notRepairedDamage'])
    new_df = new_df.append(new_row,
ignore_index=True)
    labels =
['gearbox','notRepairedDamage','model','brand','fuelT
ype','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','monthOfR
egistration'] + [x+'_labels' for x in labels]]


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

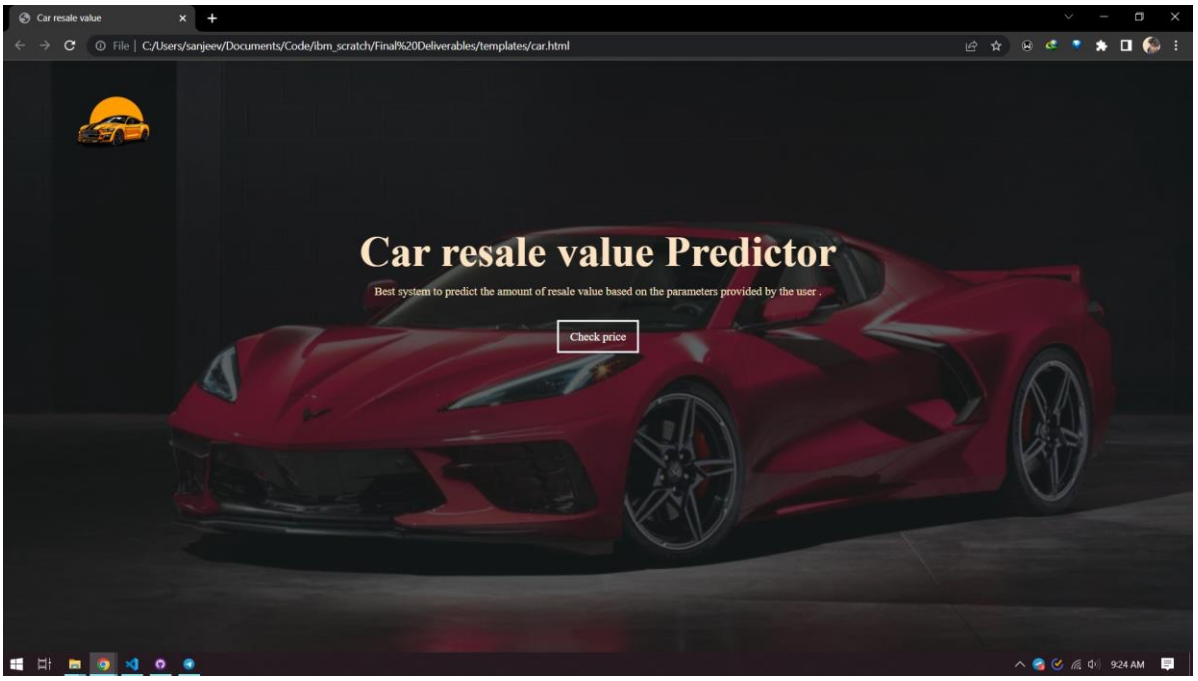
```
    #predict =
predictions['predictions'][0]['values'][0][0]
    print("Final prediction :",predict)

    return
render_template('predict.html',predict=predict)

if__name__=='__main__':
    reg_model = load_model()#load the saved model
    app.run(debug=True)
```

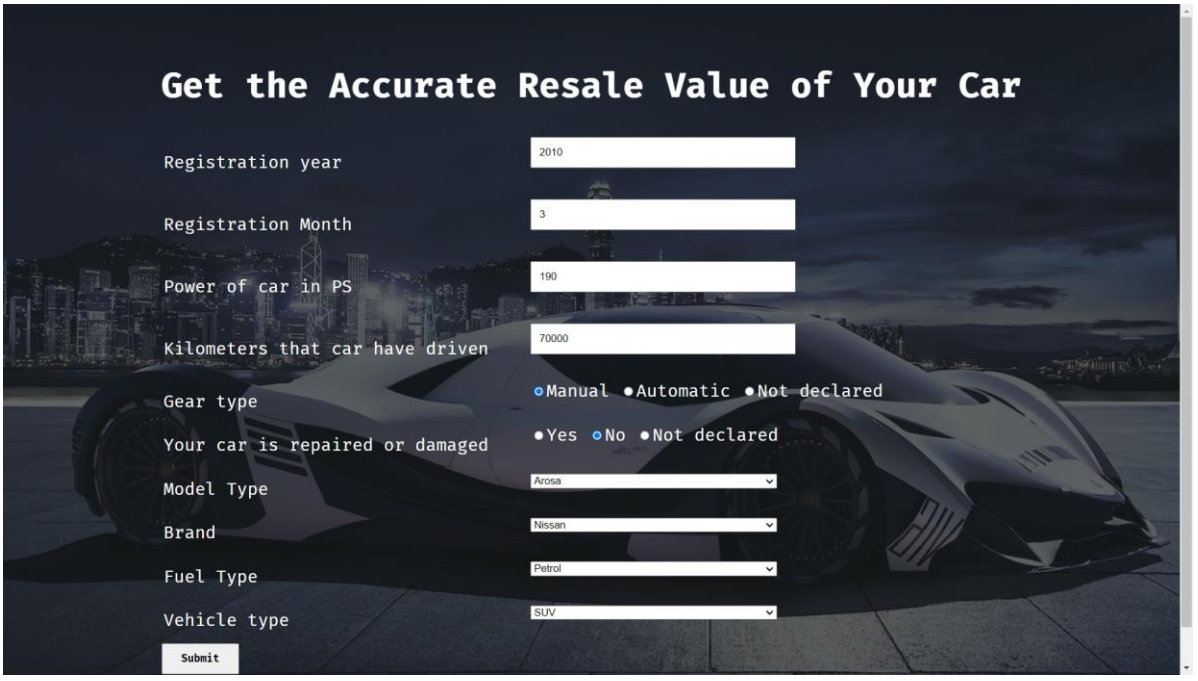
8. TESTING

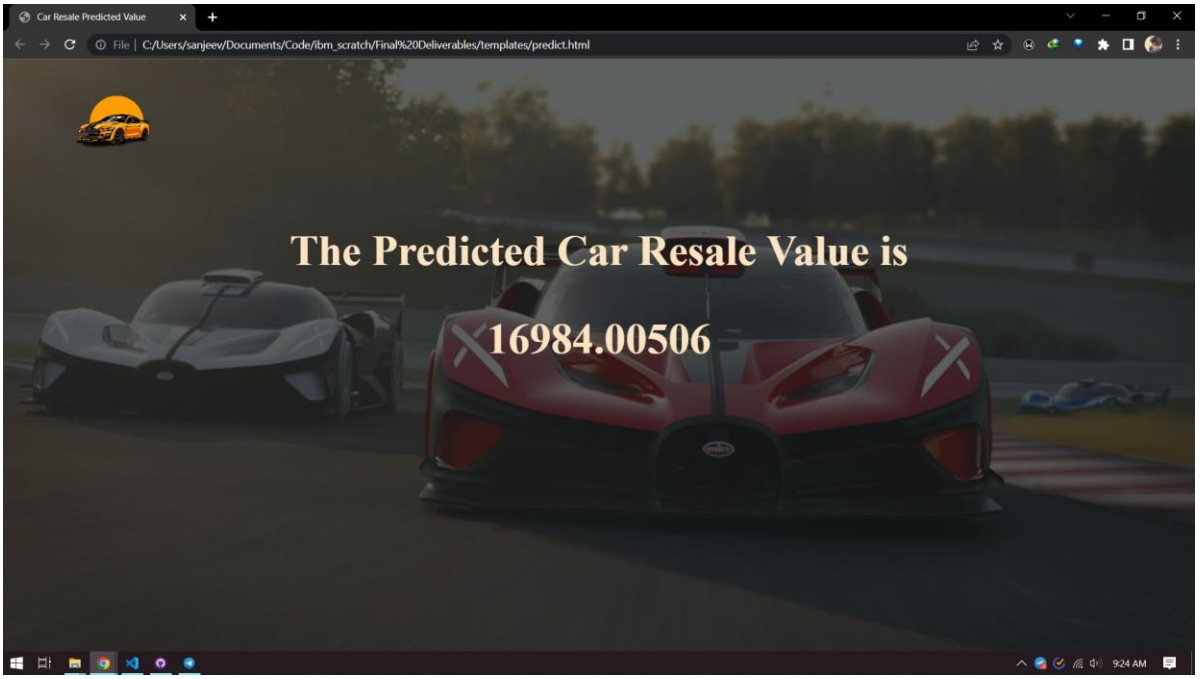
8.1. Test Cases

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

8.2. User Acceptance Testing






Car Resale Predicted Value
File | C:/Users/sanjeev/Documents/Code/ibm_scratch/Final%20Deliverables/templates/predict.html
The Predicted Car Resale Value is
16984.00506
9:24 AM

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

Pros

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

Cons

- Less effective

11. CONCLUSION

With 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 that effectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of a used car price prediction.

12. FUTURE SCOPE

In the 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 on car prices which can help to improve the 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.

13. APPENDIX

```
# App.py

# Import Libraries

import pandas as pd

import numpy as np

from flask import Flask, render_template, Response,

request

import pickle

from sklearn.preprocessing import LabelEncoder


app = Flask(__name__)#initiate flask app


def load_model(file='model.sav'):#load the saved

model

    return pickle.load(open(file, 'rb'))


@app.route('/')

def index():#main page
```

```
    return render_template('car.html')


@app.route('/predict_page')
def predict_page():#predicting 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','ge
arbox',
```

```
'powerPS','model','kilometer','monthOfRegistration','f
uelType',
        'brand','notRepairedDamage'])
    new_df = new_df.append(new_row,
ignore_index=True)
    labels =
['gearbox','notRepairedDamage','model','brand','fuelT
ype','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','monthOfR
egistration'] + [x+'_labels' for x in labels]]

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



    #predict =
predictions['predictions'][0]['values'][0][0]
    print("Final prediction :",predict)

    return
render_template('predict.html',predict=predict)
```

```
if__name__=='__main__':
    reg_model = load_model()#load the saved model
    app.run(debug=True)
```

GitHub https://github.com/IBM-EPBL/IBM-Project-55149-1666081600

VideoLink -> https://drive.google.com/file/d/1z-H56lYsm-OvTun2ytBHCw0j_BWb706o/view?usp=share_link