Car Resale Value prediction

PROJECT ID : PNT2022TMID50759

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.

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 dataset. This ability to capture data, analyses it and use it to personalize a shopping experience or implement is the future of retail.

Purpose

Car resale value prediction helps the user to predict the re sale value of the car depending upon various features like kilo-meters 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

A literature review serves to situate the current study within the body of the relevant literature and to provide context of the reader.

Sl. No.	Title	Author & Publications	Year	Description
1.	Used Car Price Prediction using K-Nearest Neighbor Based Model	K. Samruddhi, Dr. R. Ashok Kumar & *International Journal of Innovative Research in Applied Sciences and Engineering(IJIRASE).*	2020	A machine learning model is proposed to estimate the cost of the used cars using the K-Nearest Neighbor algorithm. The model is trained with used cars data for test ratios. Then the proposed model is cross-validated using fold method to examine the performance.

Sl. No.	Title	Author & Publications	Year	Description
1.	Used Car Price Prediction using K-Nearest Neighbor Based Model	K. Samruddhi, Dr. R. Ashok Kumar & *International Journal of Innovative Research in Applied Sciences and Engineering(IJIRASE).*	2020	A machine learning model is proposed to estimate the cost of the used cars using the K-Nearest Neighbor algorithm. The model is trained with used cars data for test ratios. Then the proposed model is cross-validated using fold method to examine the performance.

Sl. No.	Title	Author & Publications	Year	Description
2.	USED CAR PRICE PREDICTION AND LIFE SPAN	Aditya Nikhade, Rohan Borde & *International Advanced Research Journal in Science, Engineering and Technology.*	2021	By using the optimized BP neutral network algorithm,the price evaluation model based on big data analysis is proposed, which takes the advantage of widely circulated vehicle.

Sl. No.	Title	Author & Publications	Year	Description
3.	PREDICTIVE ANALYSIS OF USED CAR PRICES USING MACHINE LEARNING	Ashutosh Datt Sharma, Vibhor Sharma, Sahil Mittal, Gautam Jain, Sudha Narang & *International Research Journal of Modernization in Engineering Technology and Science.*	2021	Employing various Machine Learning Algorithms, a statistical model was build and based upon the given data and features set to estimate the price of used cars. After applying various regression algorithm on the model, they concluded Decision Tree algorithm was the best performer.

Sl. No.	Title	Author & Publications	Year	Description
4.	Car Resale Value Prediction System	Dhwani Nimbark, Akshat Patel, Sejal Thakkar & *International Research Journal of Engineering and Technology(IRJET).*	2021	Different algorithms had used like Support Vector Regression, Logistic Regression, Random Forest Regression and Gradient Boosting Regression for developing Car resale value prediction systems by considering different features of the car.

EXISTING SYSTEM AND PROBLEM

The existing system for used car price prediction is Cars24. Cars24 is a cutting-edge e-commerce platform for pre-owned cars, which strives to make buying and selling cars a breeze. The platform offers a wide range of certified cars along with smooth and convenient home delivery facilities with the click of a button.

Disadvantages

- Limited Market
- Difference Between Online Valuation & Post Inspection Price

- Seasonal Business

REFERENCES

1) 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). Volume 4, Issue 3, September 2020.

2) Aditya Nikhade, Rohan Borde, "USED CAR PRICE PREDICTION AND LIFE SPAN", *International Advanced Research Journal in Science, Engineering and Technology*.VoL.8, Issue 12, December 2021.

3) *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, Volume: 03,* Issue: 06, June-2021.

4) Dhwani Nimbark, Akshat Patel, Sejal Thakkar, "Car Resale Value Prediction System", International Research Journal of Engineering and Technology(IRJET), Volume: 08, Issue: 05, May 2021.

PROBLEM STATEMENT

Second hand car is predominantly used by all. So, the prediction of used car price becomes the significant and interesting area of analysis. The price of the used car depends on the factors like color, model, mileage and number of seats. Car resale value evaluator is developed to predict the price to attain benefits to buyer and seller.

1). Raju is an Auto Sales Representative who needs a way to accurately predict the value of used car because he needs to satisfies his customer.

2). User needs a way to predict the value of used car by taking it's model name and seller type because she doesn't want very old model cars and seller type is inguinal.

3). Ravi needs a way to predict the value of used car because it's difficult to anticipate how much a used car will sell for.

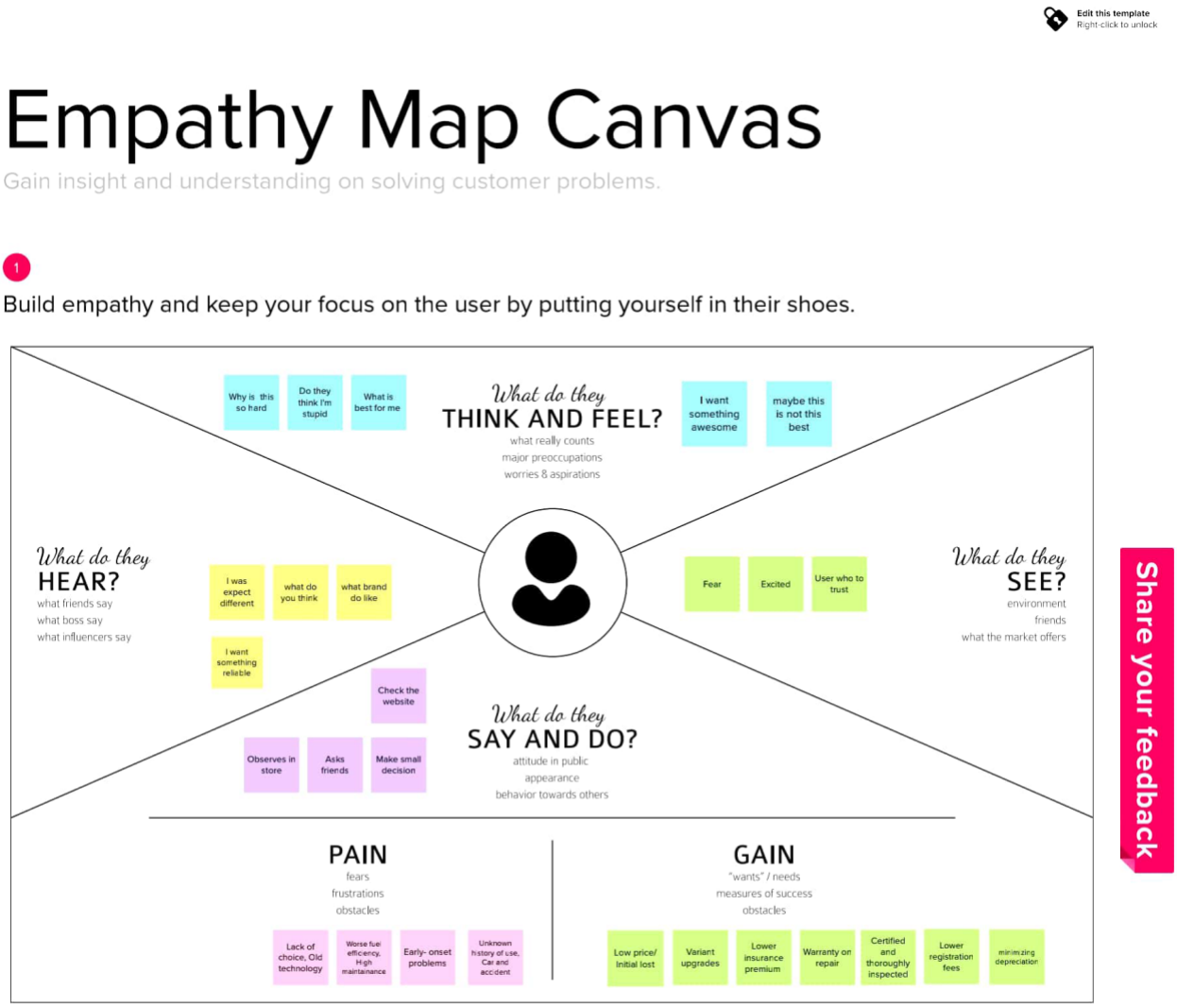
4). User is an explorer who needs a way to predict the value of used car based on mileage driven and transmission types because he wants to be low level petrol in run more kilometers and automatic types.

5). User is an owner who needs a way to predict the accurate value of used car because he wants to know the actual worth of their car and to sell it.

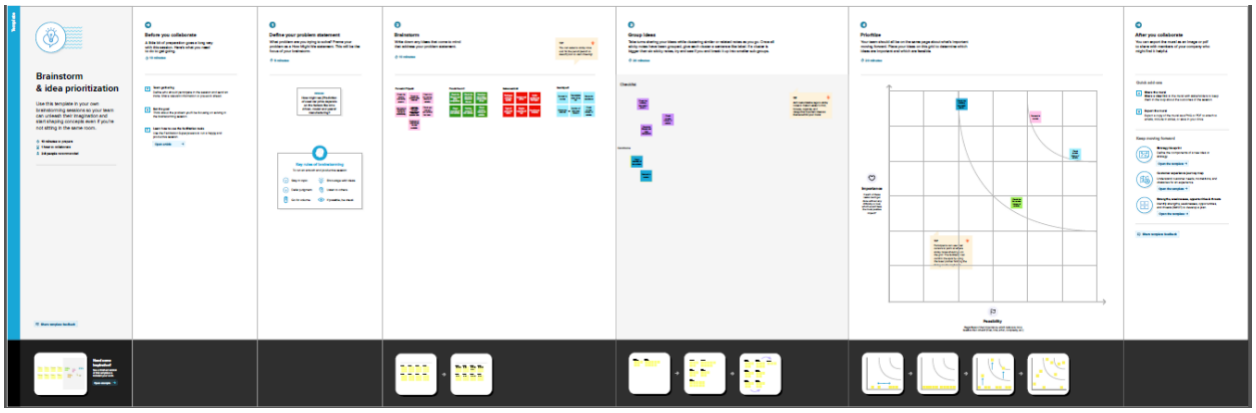
IDEATION PROPOSED SOLUTION

The project aims to develop an application and the application uses IBM cloud storage for storing data.

EMPATHY MAP



IDEATION & BRAINSTORMING

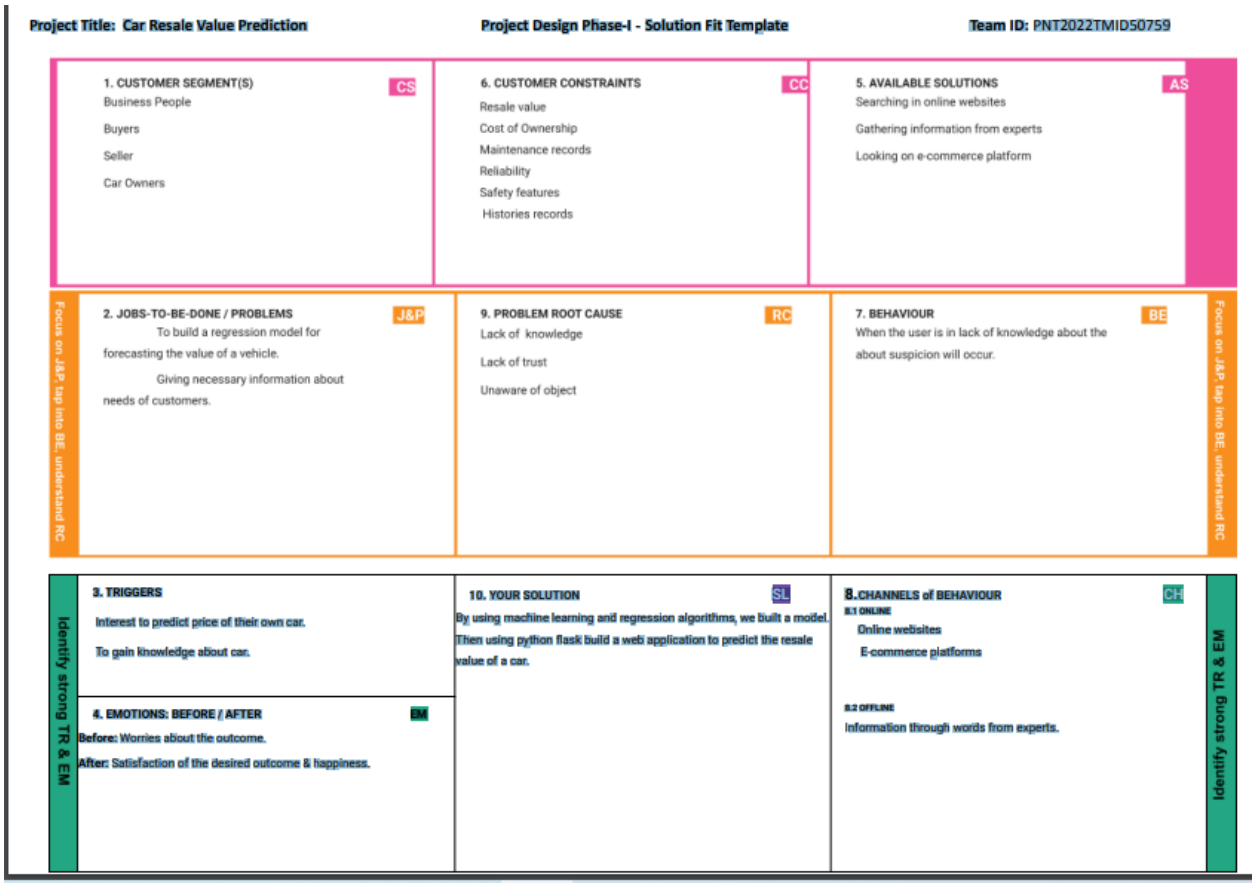


PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Second hand car is predominantly used by all. So the prediction of used car price becomes the significant and interesting area of analysis. The price of the used car depends on the factors like miles driven, model and year of manufacturing. Car resale value evaluator web application is developed to predict the price to attain benefits to buyer and seller.
2.	Idea / Solution description	The overall proposed idea is to predict the car resale value and show it to the user. This idea can be implemented and could be presented to the customer. This involves two phases. Phase I is collecting the dataset for training the car resale value prediction model. Testing the car resale value prediction model. Phase II involves creating a website for presenting the entire solution as a customized GUI this would be very useful for the user to utilize this solution the user will be asked to enter the details for prediction like model, price, design, kilometres Covered, interior look, colour. If user clicks the predict option, the predicted resale value will be displayed in the website.

3.	Novelty / Uniqueness	The unique solution in which the car maximum details are obtained from registration number of the car. Additionally details like any internal or external damages, driven kilometres based on these details, The price value is predicted it show the price more accurate.
4.	Social Impact / Customer Satisfaction	By using this model the car seller know the price of their car model. These make trust among the car seller the car. It leads to increase the sale of used car market people wishes to buy used car instead of buying new car with high cost.
5.	Business Model (Revenue Model)	Revenue mode by promoting car models and advertising car. But the project is absolutely free for users. If they interested to advertise their car model then they can approach any third party serving. Also car seller sell their car through our platform. It should be paid for the car seller to
		verify their car. Based on their provided details should be correct for this the separate service charge should be mode.
6.	Scalability of the Solution	we are using time series analysis so, When historical data for a product ss is available and patterns are obvious, organisations typically employ the time series analysis technique to demand forecasting. A time series analysis can help you detect seasonal variations in demand, cyclical patterns, and major sales trends.The time series analysis approach works best for well-established organisations with several years of data to work with and very steady trend patterns

PROPOSED SOLUTION FIT



REQUIREMENT ANALYSIS

Functional Requirements

Following are the functional requirements of the proposed solution

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	• Registration through Form • Registration through Gmail • Registration through mobile number
FR-2	User Confirmation	• Confirmation via Email • Confirmation via OTP
FR-3	Car registration	• Registering car details
FR-4	Value Prediction	• Predicting the car resale value

Non –functional Requirements

Non-functional Requirements:

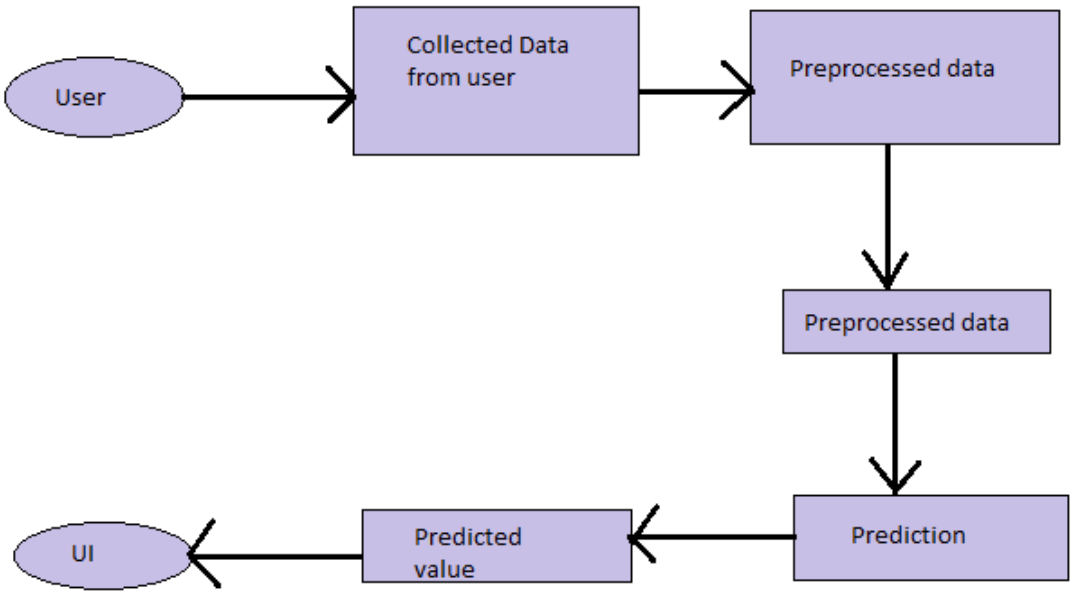
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Predicting the resale value
NFR-2	**Security**	Providing security to the website
NFR-3	**Reliability**	Providing high reliability by predicting values for different type of cars
NFR-4	**Performance**	Providing high performance by using some machine learning techniques
NFR-5	**Availability**	The webpage must not crash due to network traffic and must be available for user at anytime
NFR-6	**Scalability**	Predicting values for different types of cars

PROJECT DESIGN

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

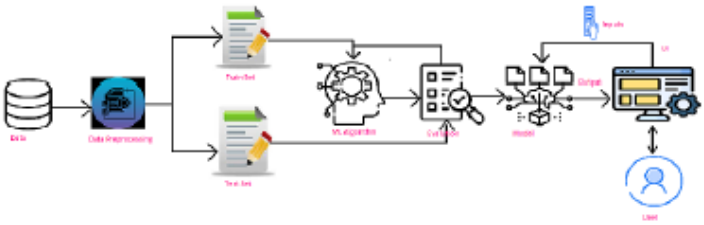


TECHNICAL & SOLUTION ARCHITECTURE

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

Reference: https://lh6.googleusercontent.com/alt4R7xSXSjqDVTmV8-_RmkIANAGMhrO35Li6HtVnUPkq1FAMu2Iq6kmTqjXRdKT45-5m0BdgT8cGIOBJKehp35DRPrTElRLHOohSKTLOMoigA12XLokcj2AdrVzQhvegUsZ8MeO



Guidelines:

1. Include all the processes (As an application logic / Technology Block)
2. Provide infrastructural demarcation (Local / Cloud)
3. Indicate external interfaces (third party API's etc.)
4. Indicate Data Storage components / services
5. Indicate interface to machine learning models (if applicable)

Solution Architecture:

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to: • Find the best tech solution to solve existing business problems. • Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders. • Define features, development phases, and solution requirements. • Provide specifications according to which the solution is defined, managed, and delivered. Example –

Solution Architecture Diagram:

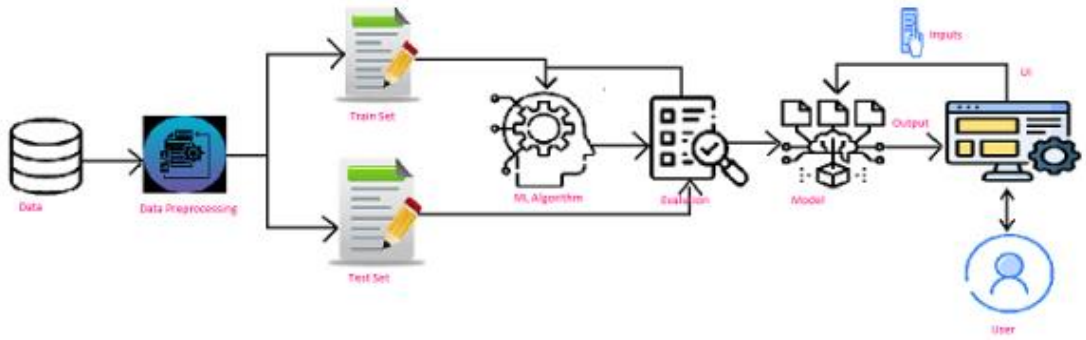


Figure 1: Architecture and data flow of the car resale application Reference: https: https://lh6.googleusercontent.com/alt4R7xSXSjqDVTmV8-_RmkIANAGMhrO35Li6HtVnUPkq1FAMu2Iq6kmTqjXRdKT45-5m0BdgT8cGIOBJKehp35DRPrTElRLHOohSKTLOMoigA12XLokcj2AdrVzQhvegUsZ8MeO.

PROJECT PLANNING & ESTIMATION

Sprint & Planning Estimation

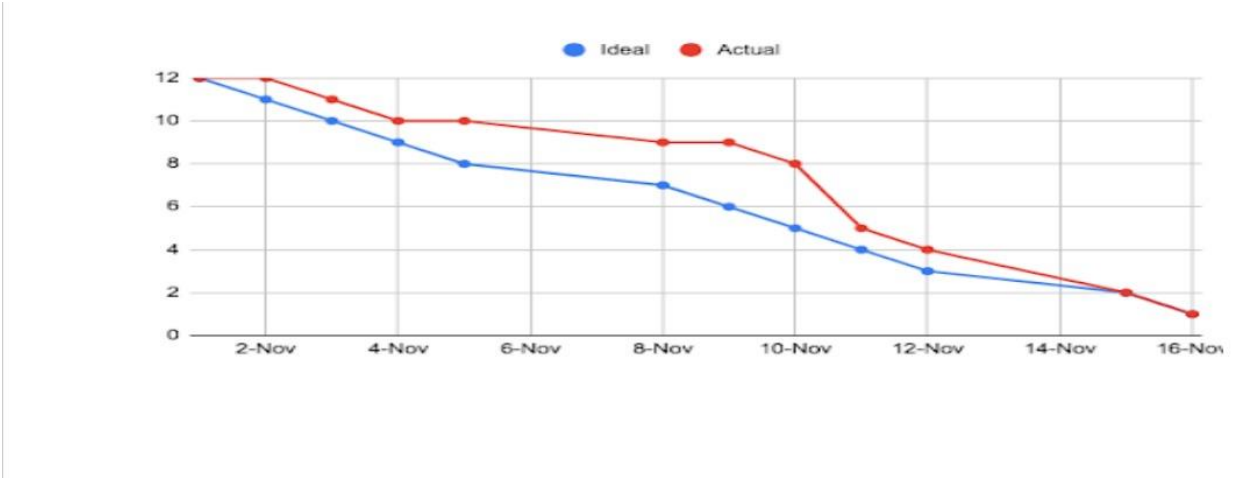
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Priority	Team Membe
Sprint-1	Pre- process the data	USN-1	Collect and download the Dataset	High	Parvathi Priya

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Priority	Team Membe
Sprint-1		USN-2	Import required libraries	High	Petchikani
Sprint-1		USN-3	Read and clean the dataset	Low	Santhiya
Sprint-2	Model Building	USN-1	Split the data into independent and dependent variables	Medium	Selvamathi
Sprint-2		USN-2	Build regression model	High	Parvathi Priya Selvamathi
Sprint-3	Application Building	USN-1	Build python application	Medium	Petchikani & Santhiya
Sprint-3		USN-2	Test the application model	High	Parvathi Priya Petchikani
Sprint-4	Train the model	USN-1	Train the model	High	Petchikani & Selvamathi

Velocity= Sprint Duration/ Velocity=20/10=2

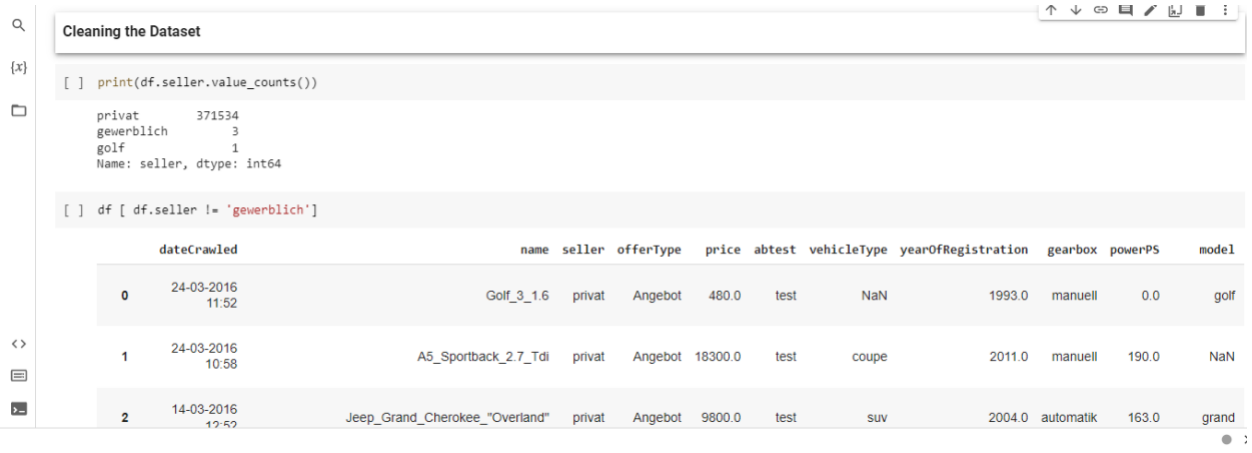
Reports From JIRA

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.

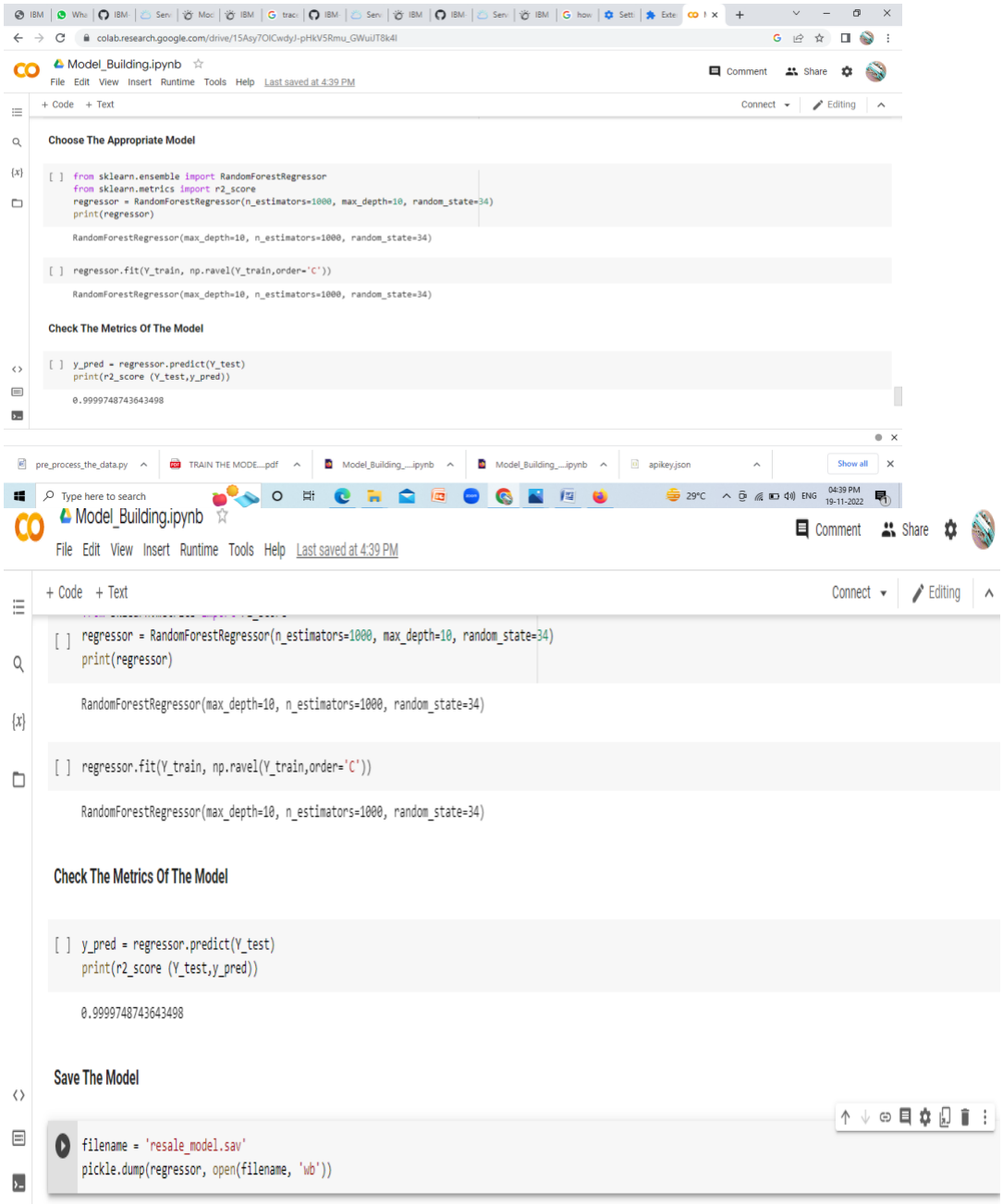


Coding & Solution

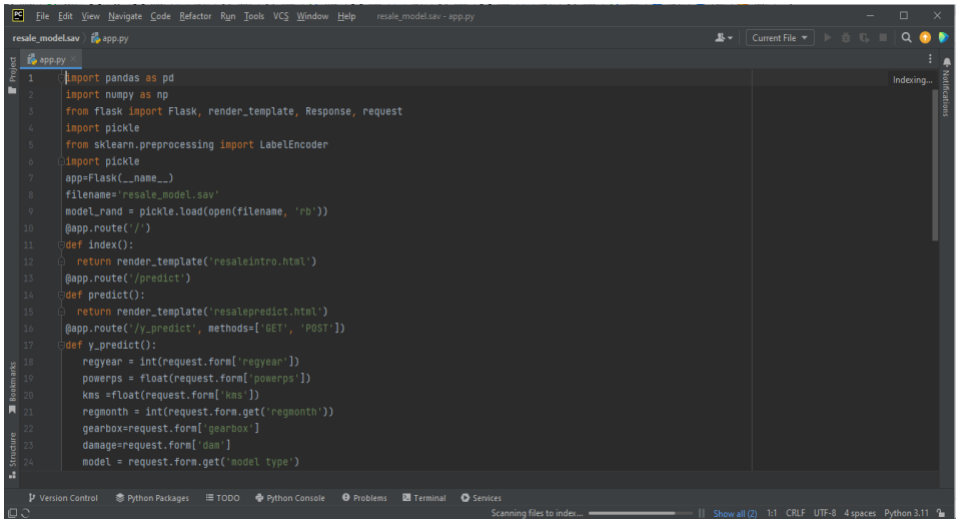
Preprocess the data



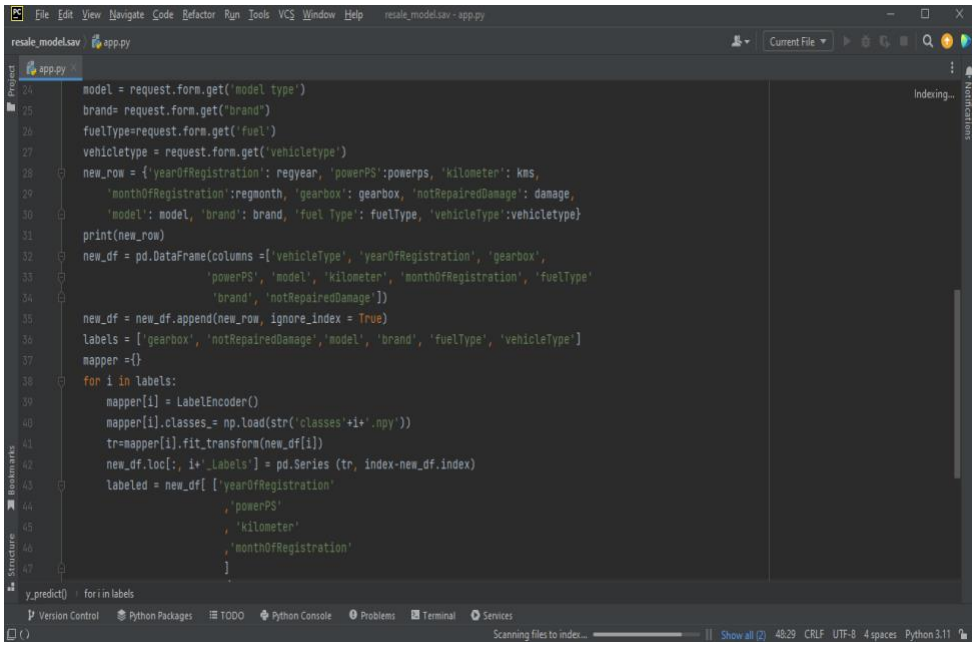
Model Building



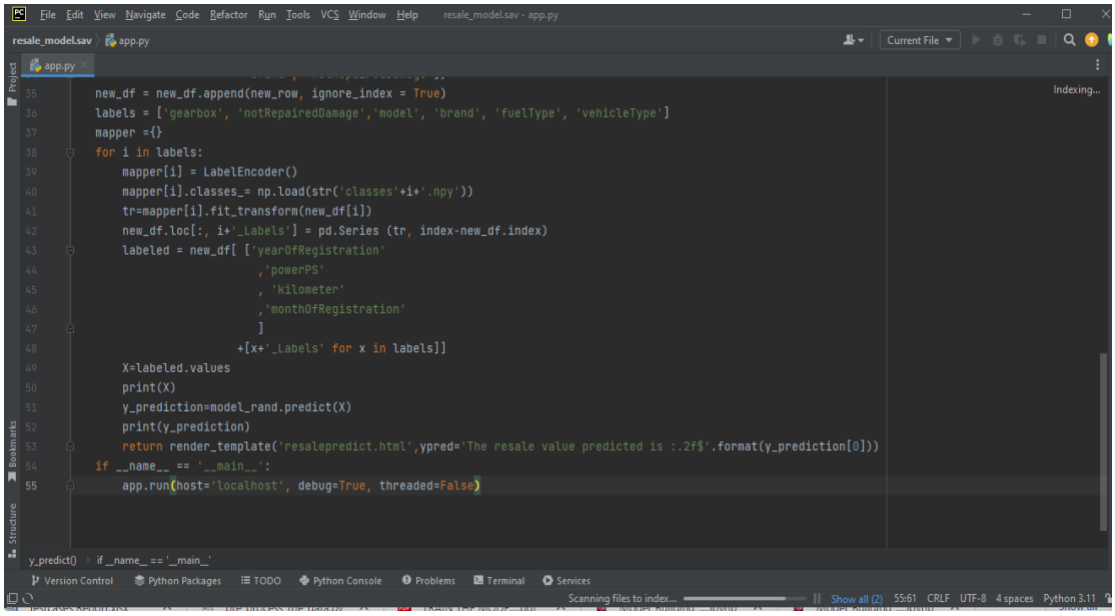
Integrating With Flask



```
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
app=Flask(__name__)
filename='resale_model.sav'
model_rand = pickle.load(open(filename, 'rb'))
@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('model type')
```



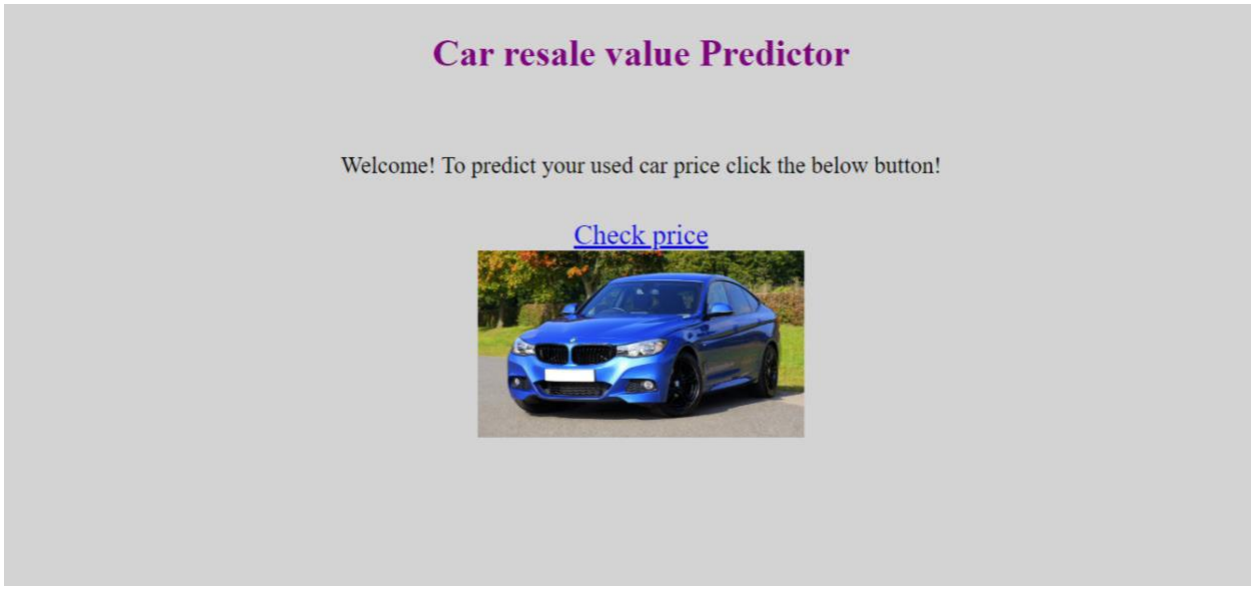
```
    model = request.form.get('model type')
    brand= request.form.get("brand")
    fuelType=request.form.get('fuel')
    vehicletype = request.form.get('vehicletype')
    new_row = {'yearOfRegistration': regyear, 'powerPS':powerps, 'kilometer': kms,
        'monthOfRegistration':regmonth, 'gearbox': gearbox, 'notRepairedDamage': damage,
        'model': model, 'brand': brand, 'fuel Type': fuelType, 'vehicleType':vehicletype}
    print(new_row)
    new_df = pd.DataFrame(columns =['vehicleType', 'yearOfRegistration', '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'))
        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'
                           ]
```



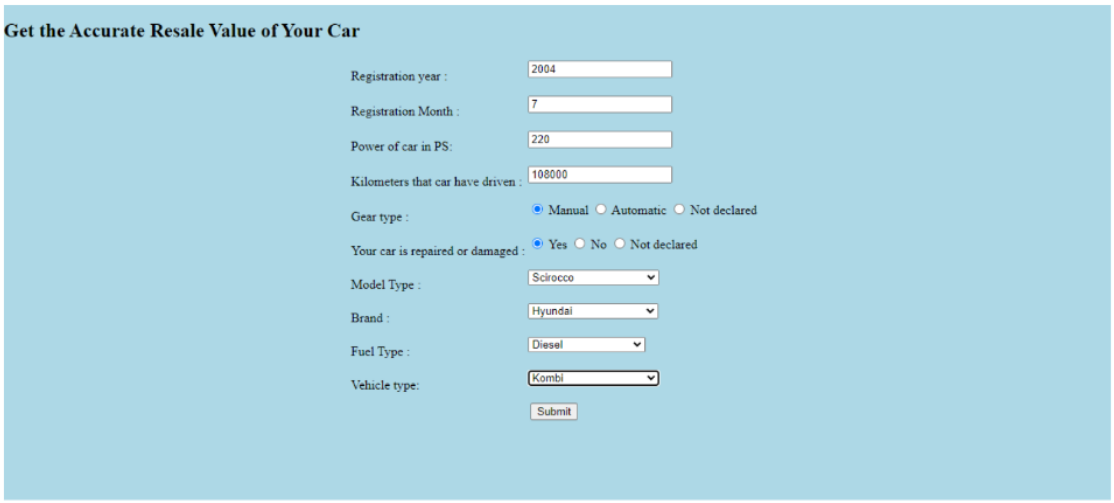
Application building

Execute & Test Model

1.Home Page



2. Car Data Entry Page



3.Display Page

Advantages

car resale value prediction helps the user to predict the resale value of the car depending upon various features like kilometres 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..

DISADVANTAGES:

Less effective.

Conclusion

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.

Appendix

GitHub Link: https://github.com/IBM-EPBL/IBM-Project-49567-1660825950

Demo Link: **https://drive.google.com/drive/folders/1mZi3E2JDHpmpEwhQL**