ROEVER ENGINEERING COLLEGE

CAR RESALE VALUE PREDICTION

NALAIYA THIRAN PROJECT REPORT
2022

Team ID: IBM-Project-42049-1660648019

College Name: Roever Engineering College

Submitted by

Mala N	**(813519104027)**
Elavalagi P	**(813519104010)**
Gowsing prabha P	**(813519104012)**
Maniyarasi M	**(813519104031)**
Narmatha S	**(813519104034)**

TABLE OF CONTENT

GitHub & Project Demo Link

1. INTRODUCTION

Determining whether the listed price of a used car is a challenging task, due to the manyfactors that drive a used vehicle's price on the market. The focus of this project is developing machine learningmodels that can accurately predict the price of a used car based on its features, in order to make informedpurchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices ofdifferent makes and model. We will compare the performance of various machine learning algorithms like LinearRegression, Ridge Regression, Lasso Regression, Elastic Net, Decision Tree Regressor and choose the best out ofit. Depending on various parameters we will determine the price of the car. Regression Algorithms are usedbecause they provide us with continuous value as an output and not a categorized value because of which it willbe possible to predict the actual price a car rather than the price range of a car. User Interface has also beendevelopedwhichacquiresinputfromanyuseranddisplaysthePriceofacaraccordingtouser'sinputs.

1.1. PROJECTOVERVIEW:

systemisdefinedinthepythonlanguagethatpredictstheamount of resale value based on the giveninformation. The system works on the trained dataset of the machine learning program that evaluates thepreciseevalueofthecar.Usercanenterdetailsonlyoffieldslikepurchasepriceofcar,kilometersdriven,fuel ofcar, year of purchase.Upon form submission, the data is sent to the ML model via Flask API and the modelresponds with a predicted resale value of the car based on user input. This prediction is displayed on the webpage using a render template. Thus, with minimal information and without human intervention or manualexamination,ausercanpredicttheresalevalueofhiscar.

1.2. PURPOSE

This resale value prediction system is made for general purpose to just predict the amount that can be roughlyacquired by the user. We try to predict the amount of resale by best 70% accuracy so the user can getestimated value before he resales the car and doesn't make a deal in lossThe main idea of making a car resalevaluepredictionsystemistogethands-onpracticeforpythonusingDataScience.Carresalevaluepredictionis the system to predict the amount of resale value based on the parameters provided by the user. User entersthe details of

the car into the form given and accordingly the car resale value is predicteCar resale valuepredictionsystemismadewiththepurposeofpredictingthecorrectvaluationofusedcarsthathelpsusers tosell the car remotely with perfect valuation and without human intervention in the process to eliminate biasedvaluation.

2. LITERATURESURVEY:

Several studies and related works have been done previously to predict used car prices around the world using different methodologies and approaches, with varying results of accuracy from 50% to 90%. In (Pudaruth, 2014) the researcher proposed to predict used car prices in Mauritius, where he applied different machine learning techniques to achieve his results like decision tree, K-nearest neighbours, Multiple Regression and Naï ve Bayes algorithms to predict the used cars prices, based on historical data gathered from the newspaper.

Achieved results ranged from accuracy of 60-70 percent, the author suggested using more sophisticated models and algorithms to make the evaluation, with the main weakness off the decision tree and naï ve Bayes that it is required to discretize the price and classify it which accrue to more inaccuracies. Moreover, he suggested a larger set of data of data to train the models hence the data gathered was not sufficient.

(Monburinon, et al., 2018) Gathered data from a German e-commerce site that totalled to 304,133 rows and 11 attributes to predict the prices of used car using different techniques and measured their results using Mean Absolute Error (MEA) to compare their results. Same training dataset and testing dataset was given to each model. Highest results achieved was by using gradient boosted regression tree with a MAE of 0.28, and MEA of 0.35 and 0.55 for mean absolute error and multiple linear regression respectively. Authors suggested adjusting the parameters in future works to yield better results, as well as using one hot encoding instead of label encoding for more realistic data interpretations on categorical data.

(Gegic, Isakovic, Keco, Masetic, &Kevric, 2019) from the International Burch University in Sarajevo, used three different machine learning techniques to predict used car prices. Using data scrapped from a local Bosnian website for used cars totalled at 797 car samples after preprocessing, and proposed using these methods: Support Vector Machine, Random Forest and Artificial Neural network. Results have shown using only one machine learning algorithm achieved results less than 50%, whereas after combing the algorithms with pre
calcification of prices using Random Forest, results with accuracies up to 87.38% was recorded.

(Noor & Jan, 2017) were able to achieve high level of accuracy using Multiple linear regression models to predict the price of cars collected from used cars website in Pakistan called Pak Wheels that totalled to 1699 records after preprocessing, and where able to achieve accuracy of 98%, this was done after reducing the total amount of attributes using variable selection technique to include significant attributes only and to reduce the complexity of the model.

(K.Samruddhi& Kumar, 2020) Proposed using Supervised machine leaning model using K-Nearest Neighbour to predict used car prices from a data set obtained from Kaggle containing 14 different attributes, using this method accuracy reached up to 85% after different values of K as well as Changing the percent of training data to testing data, expectedly when increasing the percent of data that is tested better accuracy results are achieved. The model was also cross validated with 5 and 10 folds by using K fold method.

(Gongqi, Yansong, &Qiang, 2011) proposed using Artificial Neural Network (ANN) through a combined method of BP neural network and nonlinear curve fit and have achieved accurate value prediction with a feasible model.

(Listiani, 2009) used Support Vector Machines to evaluate leased cars prices, results have shown that

SVM is far more accurate in large dataset with high dimensional data than Multiple linear regression. Whereas the computation Multiple linear regression can take several minutes and the SVM would take up to a day to compute the results. Multiple linear regression may be simple, but SVM is far more accurate. Moreover, the study includes Samples with up to 178 attributes which is far more than the proposed variable in our study, hence the use of multiple linear regression may be more suitable in our case.

(Kuiper, 2008) Collected data from General Motor of cars that are produced in 2005, where he as well used variable selection technique to include the most relevant attributes in his model to reduce the complexity of the data. He proposed used Multivariate regression model that would be more suitable for values with numeric format.

In order to predict the price of used cars, researchers (Nabarun Pal, 2018) used a supervised learning method known as Random Forest. Kaggle's dataset was used as a basis for predicting used car prices. In order to determine the price

impact of each feature, careful exploratory data analysis was performed. 500 Decision Trees were trained with Random Forests. It is most commonly used for classification, but they turned it into a regression model by transforming the problem into an equivalent regression problem. Using experimental results, it was found that training accuracy was 95.82%, and testing accuracy was 83.63%. By selecting the most correlated features, the model can accurately predict the car price.

In light of the number of works that have been done in this field, another group of researchers (Jian Da Wu, 2017) conducted research on this topic and tried to develop a system that consists of three components: a data acquisition system, a price forecasting algorithm, and a performance analysis. Due to its adaptive learning capability, a conventional artificial neural network (ANN) with a backpropagation network is compared to the proposed ANFIS. In the ANFIS, qualitative fuzzy logic approximation as well as adaptive neural network capabilities are included. Using ANFIS as an expert system in predicting used car prices showed better results in the experiment. Using GUI, the consumer can get accurate and convenient information about used cars' purchasing prices, and experiments proved that the proposed system could provide accurate and convenient price forecasting.

Hence, from all literature review it is concluded that used cars price prediction is an important topic which is the area of many researchers nowadays. So far, the best achieved accuracy is 83.63% on kaggle’s dataset using random forest technique. The researchers have tested multiple regressors and final model

2.1. EXISTINGPROBLEM:

The forecasts of vehicle cost from the chronicled information that has been gatheredfrom every day papers. They have utilized the administered AI strategies for foreseeing the cost of vehicles.Numerous different calculations like various straight relapse, k-closest neighbor calculations, gullible based,and some choice tree calculations additionally been utilized. Every one of the four calculations are looked atand tracked down the best calculation for forecast. They have confronted a few challenges in looking at thecalculations, by one way or another they have overseen. As indicated by creators Pattabiraman, this paper ismore focused

on the connection among vender and purchaser. To foresee the cost of four wheelers, morehighlights are required like previously given value, mileage, make, model, trim, type, chamber, liter, entryways,voyage,sound,cowhide.Utilizingthesehighlightsthecostofvehiclehasbeenanticipatedwiththeassistanceof factual investigation framework for exploratory information examination. As per creators EnisGegic et al, inthis paper the chiefly focus on gathering different information from web entryway by utilizing web scrapmethods.Furthermore,thosehavebeencontrastedandtheassistanceofvariousAIcalculation.

2.2.REFERENCES:

[1] KanwalNoor,2017,VehiclePricePredictionSystemusingMachineLearningTechniquesInternational JournalofComputerApplications.Volume167-Number9

[2] Mariana Lusitania et al, (2009). Support vector regression analysis for price prediction in a vehicle leasingapplication

[3] Richardson,M.S.(2009).Determinantsofusedvehicleresalevalue.

[4] Listiani,M.(2009).Supportvectorregressionanalysisforpricepredictioninacarleasingapplication(Doctoraldissertation,Masterthesis,TUHamburg-Harburg).

[5] Richardson, M. S. (2009). Determinants of used car resale value. Retrieved from:https://digitalcc.coloradocollege.edu/islandora/object

[6] Pudaruth,S.,2014."PredictingthePriceofUsedCarsusingMachineLearningTechniques."Vol4,Number7(2014),pp.753-76.

[7] Gokce,E.(2020,January10)."Predictingusedcarpriceswithmachinelearningtechniques."

2.3.PROBLEMSTATEMENTDEFINITION:

Defining The Problem Statements:

- **The main aim of this project is to predict the price of used cars using different Machine Learning models**

currently,if anyone wants to sell their car either they have to take their car either they have to take their car to a respective company workshop to get an estimate of the price

- **This will save customers time and help the company to reduce its cost And also streamline the process of selling used cars**

3. IDEATION &PROPOSEDSOLUTION:

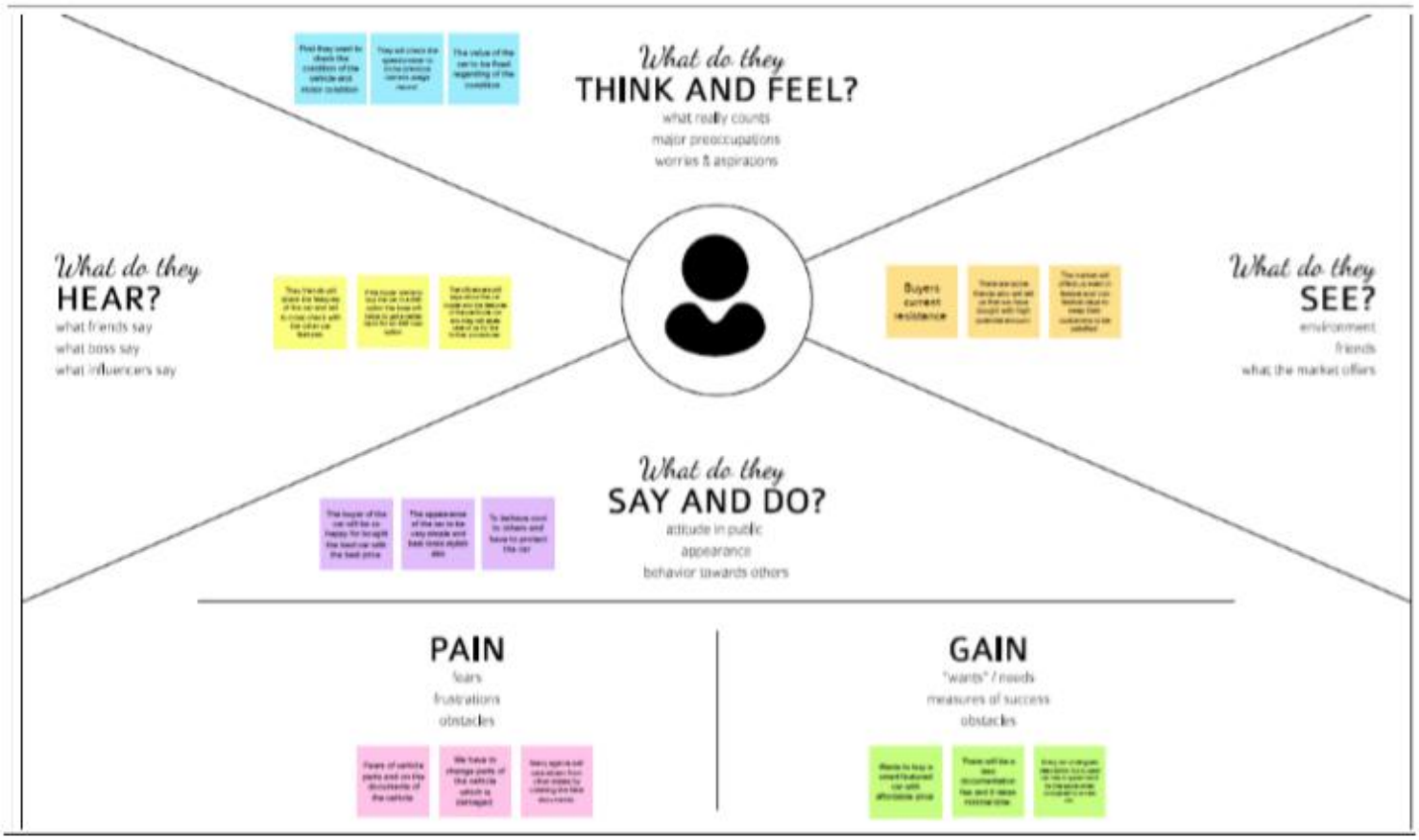
3.1. EMPATHYMAPCANVAS:

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviours and attitudes.

It is a useful tool to helps teams better understand their users. Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the

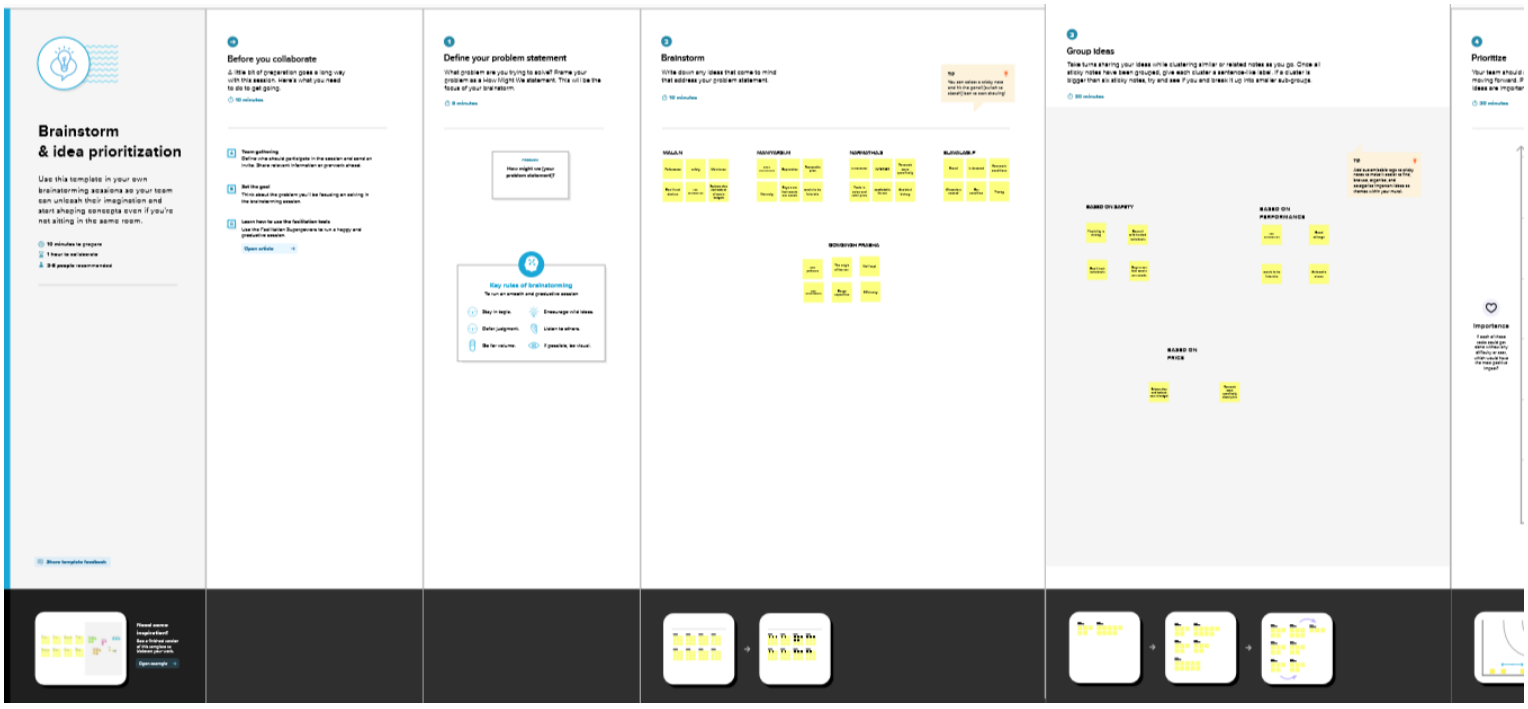
map helps participants consider things from the user's perspective along with his or her goals and challenges.

EMPATHY MAP – CAR RESALE VALUE PREDICTION:



3.2. IDEATION&BRAINSTORMING:

Brainstorming provides a free and open environment that encourages everyone within a team to participate In the creative thinking process that lead stop ro blem solving. Prioritizing volume over value, out-of-the-boxid easare welcome and builtupon, and all participant sareen courage to collaborate, helping each other develop parich amoun to fcreative solutions



3.3. PROPOSEDSOLUTION:

Project team shall fill the following information in proposed solution template.

S.No. Parameter Description

1. Problem Statement (Problem to be solved)

User will try to predict the price of used cars based on their features. As it would help the people to decide whether the used cars is worth the posted by different online used-car sites. It would also help people when they plan selling their cars! 2. Idea / Solution description We utilized a hybrid CNN-LSTM model for the task of price prediction which achieved a better performance in comparison with the baseline model. This proposed method utilizes a deep neural network involving long short-term memory (LSTM) and convolutional neural network architectures for price prediction. This system can be effective in filling such gaps which enables the users to predict the price of vehicles according to market value.

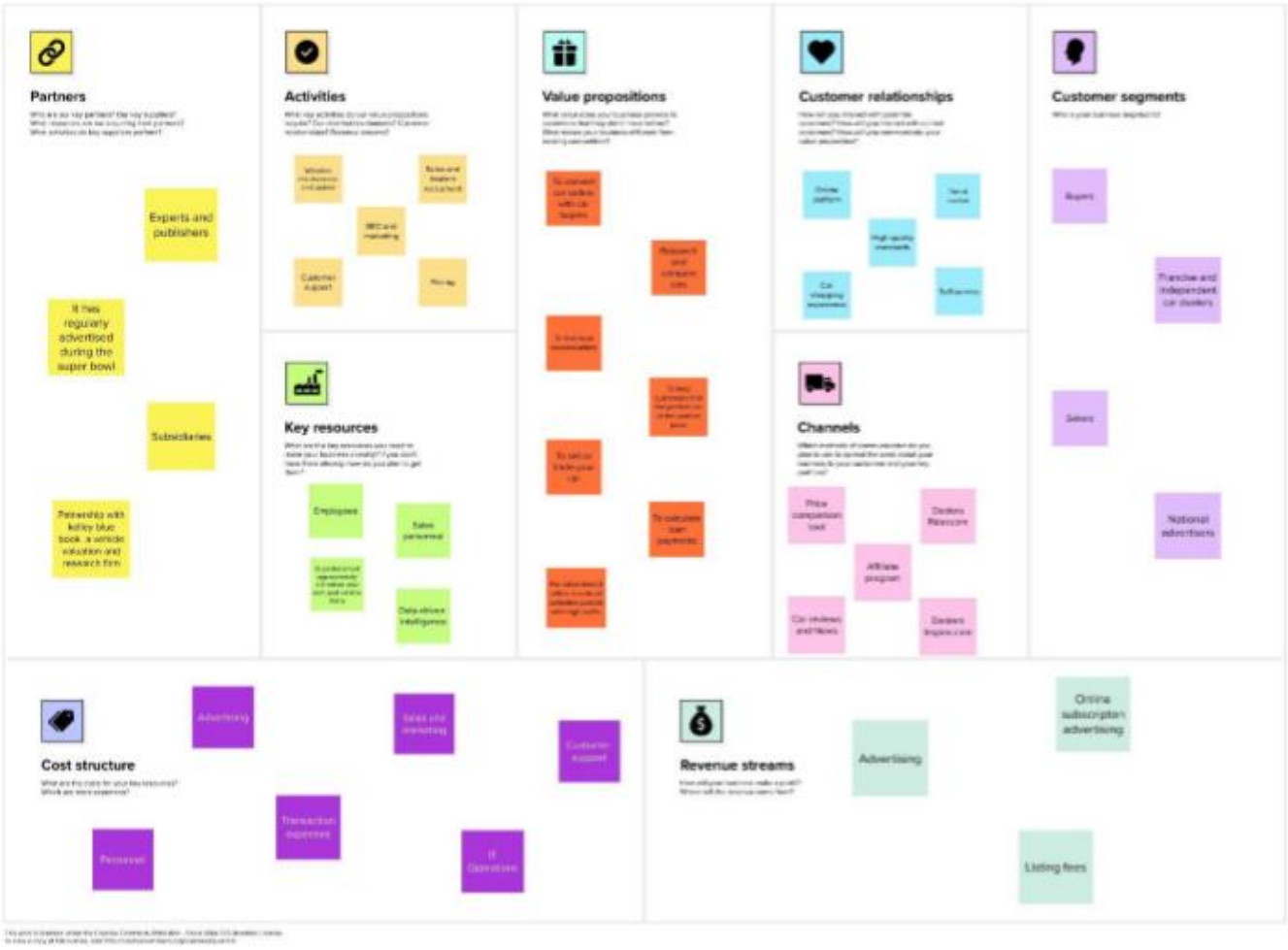
3. Novelty / Uniqueness Accuracy in price prediction

4. Social Impact / Customer Satisfaction

Offer a seamless flexible buying experience complement the in person purchasing experience by incorporating automated platforms that provide information and option to help buyers along their decision-making journey. The customer analysis section of your car dealership business plan must detail the customers you serve and/or expect to the serve. The purpose of the system is to predict the price of the used cars according to the market. . 5. Business Model (Revenue Model) Business model is attached below.

6. Scalability of the Solution There are various topics on which the prediction can be applied. Positive correlation basically relates to the concept of direct proportion whereas Negative correlation relates to the concept of inverse proportion. These estimates become the building blocks for our next step. The R2 score of Regression analysis was good for predictions and close to the original selling prices in the market. The pre-processing is required to increases the performance of UCPAS. The proposed model highlights the feasibility of combining images and textual data to make a prediction

Business model for car resale value prediction:



3.4. PROBLEMSOLUTIONFIT:

Problem – Solution Fit Template:

The Problem-Solution Fit simply means that you have found a problem with your customer and that the solution you have realized for it actually solves the customer's problem. It helps entrepreneurs, marketers and corporate innovators identify behavioral patterns and recognize what would work and why

Purpose:

❑ Solve complex problems in a way that fits the state of your customers.❑ Succeed faster and increase your solution adoption by tapping into existing mediums and channels of behavior.❑ Sharpen your communication and marketing strategy with the right triggers and messaging. ❑ Increase touch-points with your company by finding the right problem-behavior fit and building trust

by solving frequent annoyances, or urgent or costly problems.❑ Understand the existing situation in order to improve it for your target group.

Template:

Project Title: **Car Resale value Prediction**

Project Design Phase-I - Solution Fit

Team ID: PNT2022TMID40127

4. REQUIREMENTANALYSIS:

4.1. FUNCTIONALREQUIREMENT:

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 to the related websites	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Users Profile	Personal details, Bank account ,Is He/She interested in buying a car
FR-4	Gather information about the vehicle	Through the registered websites they collect information
FR-5	Display the functionality of the vehicle	Details: Fuel type , Manufactured year , Miles Driven , Record

Non-functional Requirements:

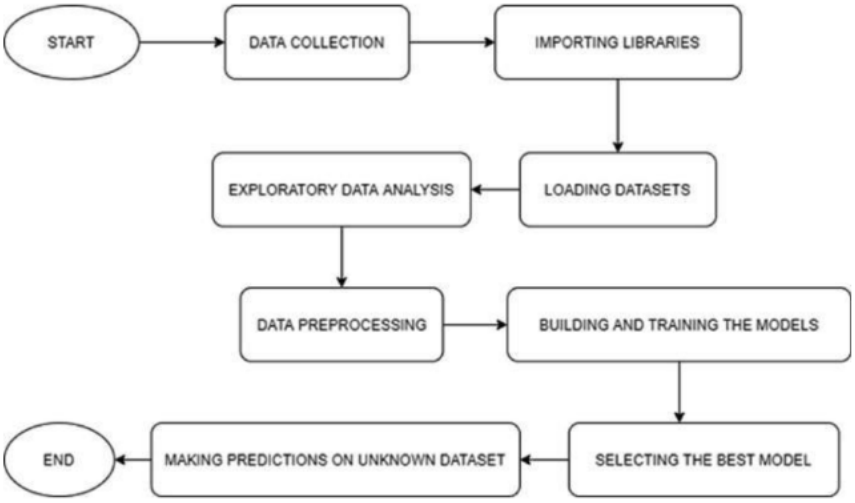
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	User friendly UI Simple and easy to Understand
NFR-2	**Security**	Aware of scams
NFR-3	**Reliability**	The system must perform without failure
NFR-4	**Performance**	The landing page must support several users must provide 5 second or less response time
NFR-5	**Availability**	Uninterrupted services must be available all time except the time of server updation.
NFR-6	**Scalability**	that can handle any amount of data and perform many computations in a cost-effective and time-saving way to instantly serve millions of users residing at global locations.

5. PROJECTDESIGN:

5.1DATAFLOWDIAGRAMS:

the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data 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 is stored.

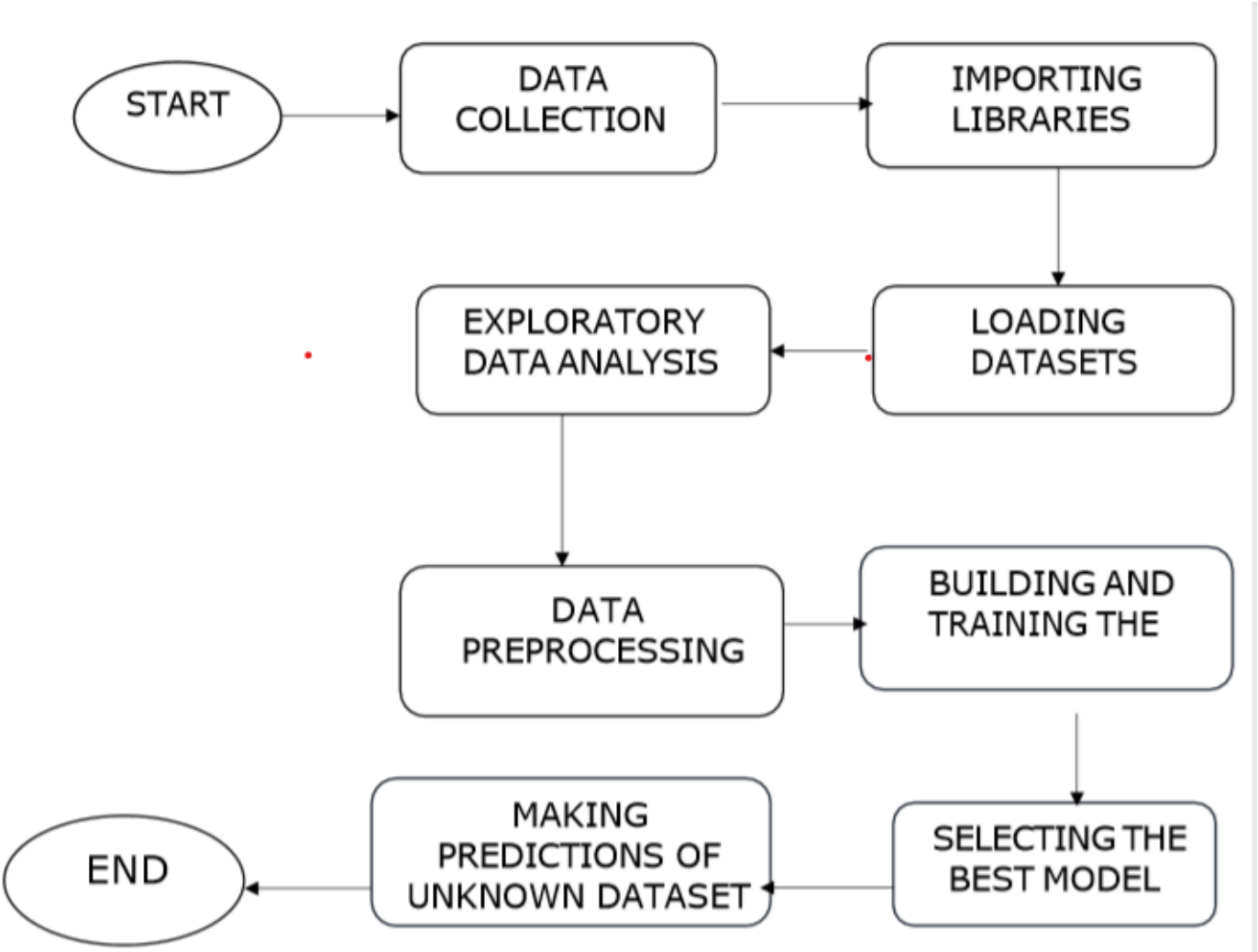


5.2. SOLUTION&TECHNICALARCHIECTURE:

solutionarchitecture:

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 solution to solve existing business problems.

Describe the structure, characteristics, behaviour , 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. Solution



Technicalarchitecture:

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

Example: Order processing during pandemics for offline mode

Reference: https://developer.ibm.com/patterns/ai-powere

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript / Angular Js / React Js etc.
2	Data preprocessing	Image of the particular vehicle uploaded through the websites and pre-processed using Machine learning algorithm	Using the various model used to process the data
3	Value prediction	Machine learning model to predict the Value of the vehicle uploaded in the website	Various models
4	Vehicle recommendation	After predicting the value , vehicle is suggested	Python
5	Database	Data's are stored in database	MySQL, NoSQL, etc.
6	Cloud Database	The model is described in the application	IBM DB2, IBM Cloudant etc.
7	File Storage	Machine learning models are used for image pre-processing, value prediction and vehicle recommendation	Data pre-processing model ,value prediction model
	External API-1	Its used for the data pre-processing	IBM server , Google drive
8	External API-2	For the users knowing value of the vehicle	Application
9	Machine Learning Model	Machine Learning Model for processing the data and predicting the value	Object Recognition Model, etc.

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1	Open-Source Frameworks	Google colaboratory , Anaconda Navigator, Jupyter Network,python flask	Data storage in google drive
2	Security Implementations	The scalability architecture is 2-tier .The client is the user and server is the IBM cloud server	SHA-256, Encryptions, IAM Controls, OWASP etc.
3	Scalable Architecture	It must support higher workloads without any issues	Models , IBM cloud
4	Availability	Availability of applications for use of load balancers, distributed servers	IBM cloud
5	Performance	Performance of the application should be high	IBM cloud

References:

https://c4model.com/

https://developer.ibm.com/patterns/online-order-processing-system-during-pandemic/

https://www.ibm.com/cloud/architecture

https://aws.amazon.com/architecture

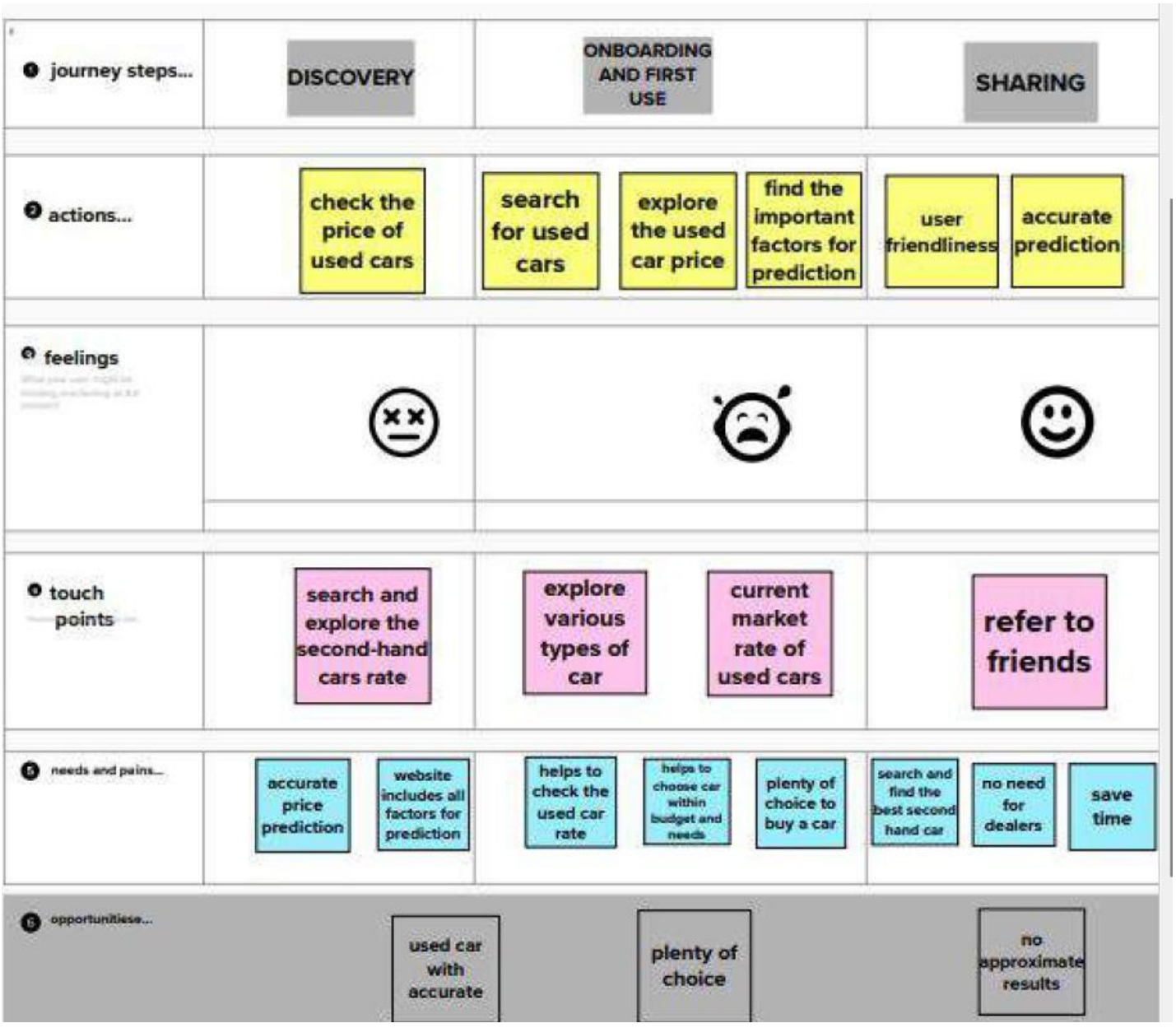
https://medium.com/the-internal-startup/how-to-draw-useful-technical-architecture-diagrams-2d20c9fda90d


Real-time Events
Clean & Prepare
App DB Cluster
Cloud SQL
Batch ETL
Cloud Dataflow
Pricing Dataset
BigQuery
ML Cluster
Kubernetes Engine
Images & Analytics generated data
Cloud Storage

5.3. USERSTORY:

Auserstoryisaninformal,generalexplanationofasoftwarefeaturewrittenfromtheperspectiveoftheenduser.Itspurposeistoarticulatehowasoftwarefeaturewillprovidevaluetothecustomer.Auserstoryisthesmallestunitofworkinanagileframework.It'sanendgoal,notafeature,expressedfromthesoftwareuser'sperspective.

A user story is an informal, general explanation of a software feature written from theperspectiveoftheenduserorcustomer.Thepurposeofauserstoryistoarticulatehowapieceof work will deliver a particular value back to the customer. The user story for the project is asfollows:



6. PROJECTPLANNING&SCHEDULING:

6.1. SprintPlanningandEstimation:

Sprintplanningisaneventinscrumthatkicksoffthesprint.Thepurposeofsprintplanningistodefinewhatcanbedeliveredinthesprintandhowthatworkwillbeachieved.Sprintplanningisdone in collaboration with the whole scrum team. In scrum, the sprint is a set period of timewhere all the work is done. However, before you can leap into action you have to set up thesprint. You need to decide on how long the time box is going to be, the sprint goal, and whereyou're going to start. The sprint planning session kicks off the sprint by setting the agenda andfocus.Ifdonecorrectly,italsocreatesanenvironmentwheretheteamismotivated,challenged,andcanbesuccessful.Badsprintplanscanderailtheteambysettingunrealisticexpectations.Thefollowingisthesprintplanningandestimationfortheproject.

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Sprint-1	Pre-process data	USN-1	Collect Dataset	1	Low
Sprint-1		USN-2	Import required libraries	1	Low
Sprint-1		USN-3	Read and clean data sets	2	Low
Sprint-2	Model building	USN-1	Split data into independent and dependent variables	3	Medium
Sprint-2		USN-2	Apply using regression model	3	Medium
Sprint-3	Application building	USN-1	Build python flask application and HTML page	5	High
Sprint-3		USN-2	Execute and test	5	High
Sprint-4	Training the model	USN-1	Train machine learning model	5	High
Sprint-4		USN-2	Integrate flask	5	High

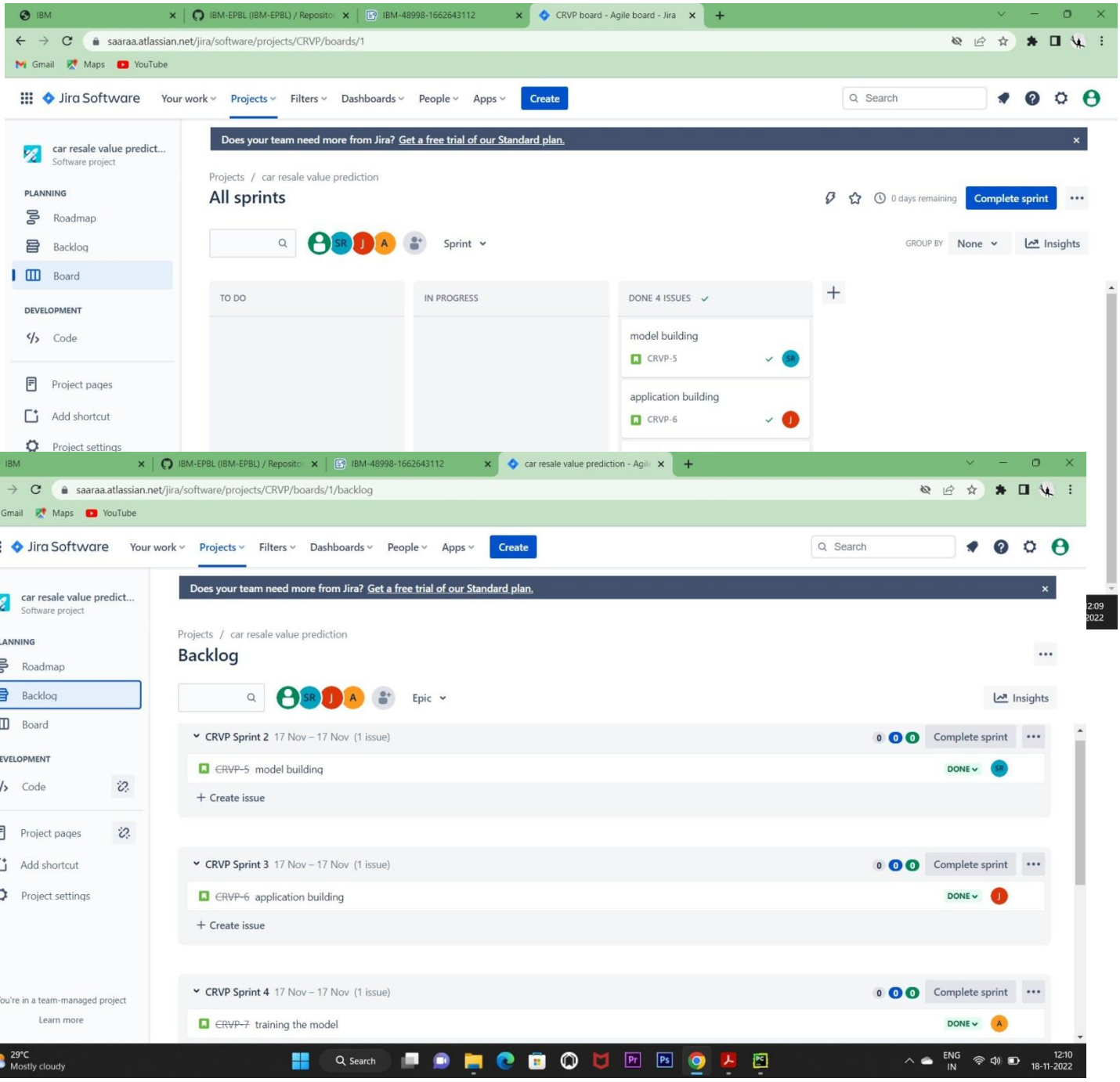
6.2SprintDeliverySchedule:

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Dataset Reading and Preprocessing	USN-1	Cleaning the dataset end splitting to dependent and independent variables	2	High	N.MALA
Sprint-2	Building the Model	USN-2	Choosing the appropriate model for building and savingthe models pickle file	1	High	M.MANIYARASI
Sprint-3	Application Building	USN-3	Using flask deploying the ML model	2	Medium	P.GOWSING HPRABHA
Sprint-4	Train the Model in ibm	USN-4	Finally train the model on IBM cloud and deploy the application	2	Medium	S.NARMATHA P.ELAVALAGI

6.3.ReportsfromJIRA:



7. CODING&SOLUTIONING:

7.1. Feature1:

Thisisasupervisedlearningproblemandcanbesolvedusingregressiontechniques.Weneedt opredictthesellingpriceofacarbasedonthegivencar'sfeatures.SupervisedRegression problemsrequirelabeleddatawhereourtargetordependentvariableisthesellingpriceofacar.Allot herfeaturesareindependentvariables.
LinearModelsarerelativelylesscomplexandexplainable,butlinearmodelsperformpoorlyondata containingtheoutliers.Linearmodelsfailtoperformwellonnon-lineardatasets.Insuchcases,non-linearregressionalgorithmsRandomForestRegressorperformbetterinfittingthenonlineardata.
Thismodelwashencechosentoaccountforthelargenumberoffeaturesinthedatasetandcompareab aggingtechniquewiththefollowinggradientboostingmethods

7.2. Feature2:

GiventheevaluationparameterstheRandomForestRegressoroutperformedasithasthehighest accuracyaswellasthelowesterrorinallthreevaluationparameters.
Asaresultofpreprocessingandtransformation,RandomForestRegressorcameoutonwit h90%accuracy

8.TESTING:

8.1TESTCASES:

• Missingvalues
ThetrainedMLmodelrequiresfewfeatureinputsforpredictingtheoutput.Failingwhich,t hemodel throws invalid Input error. All the fields in the html form have been marked requiredusingCSSandthustheusermustinputallfields.
• InvalidInput
The trained ML model requires only numerical input for all features. Thus, if the user uses symbols such as acomm a while inputting, the model may throw an error. To over come thesa me, preprocessing script is deployed in the backend which removes all unwanted characters like comma,white spaces etc. so that model get srequired input.

8.2.USERACCEPTANCETESTING

Acceptance testing focuses even more on the overall system features and functionality that are visible to the customer. Acceptance testing is of ten performed by customers to ensure customer

usability and satisfaction. The purpose of this is to briefly explain the test coverage and openissuesofthe[ProductName]projectatthetimeofthereleasetoUserAcceptanceTesting(UAT).

1. DefectAnalysis

Thisreportshowsthenumberofresolvedorclosedbugsateachseverityl evel,andhowtheywereresolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	19
Duplicate	1	0	1	0	2
External	2	1	0	2	5
Fixed	13	2	2	20	37
Not Reproduced	0	0	1	1	2
Skipped	0	0	1	1	2
Won't Fix	0	2	1	1	4
Totals	26	9	8	28	71

2. Testcaseanalysis

Thisreportshowsthenumberoftestcasesthathavepassed,failed,anduntested

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	5	0	0	5
Client Application	13	0	0	13
Security	2	0	0	2
Outsource Shipping	4	0	0	4
Exception Reporting	6	0	0	6
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. RESULTS:

9.1. PerformanceMetrics:

Performancemetricsareusedtotrackprogress.Metricsgivesomesortofconcreteanswer whicheasilycanbefollowedup.Therearedifferenttypesofmetricsusedfortesting. Theregressionmodelcanbeevaluatedonfollowingparameters:

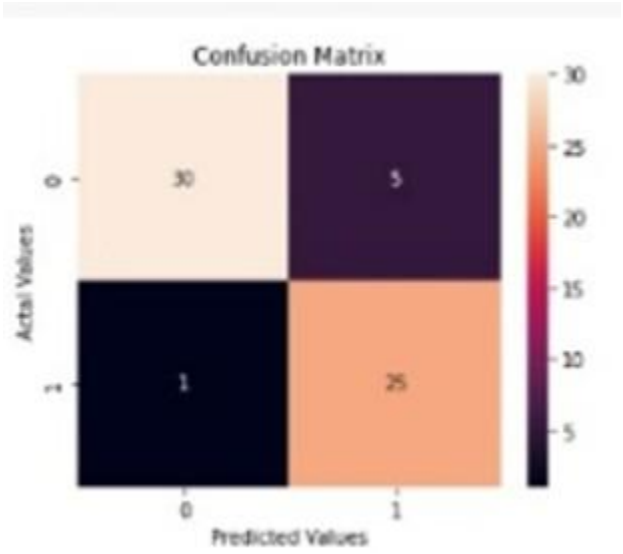
RegressionModel:

1. MeanSquareError(MSE):MSEisthesinglevaluethatprovidesinformationaboutgoodness of regression line. Smaller the MSE value, better the fit because smaller value implies smallermagnitudeoferrors.
2. RootMeanSquareError(RMSE):RMSEisthequadraticscoringrulethatalsomeasuresthea verage magnitude of the error. It is the square root of average squared difference betweenpredictionandactualobservation.
3. MeanAbsoluteError(MAE):Thismeasurerepresentstheaverageabsolutedifference betweentheactualandpredictedvaluesinthedataset.Itrepresentstheaverageresidualfromthe dataset.

```
RMSE:
0.31362502409359

MSE:
0.31362502409359

MAE:
0.09836065573770492

R2 SCORCE:
0.5978021978021978
```

ClassificationModel:

ConfusionMatrix,AccurayScore-0.9016&ClassificationReport



Classification report

```
[60] from sklearn.metrics import classification_report
     print(classification_report(original_classes, pred_classes))
```

	precision	recall	f1-score	support
0.0	0.97	0.86	0.91	35
1.0	0.83	0.96	0.89	26
accuracy			0.90	61
macro avg	0.90	0.91	0.90	61
weighted avg	0.91	0.90	0.90	61

10. ADVANTAGES&DISADVANTAGES:

Advantages:

- Accuracyofourmodelis90%.
- Predictionrunsfordifferenttypesofcars.

Disadvantages:

- Accuracycanbeimproved.
- Predictionisdoneusingonlyafewcriteria.

11. CONCLUSION:

Westartedwithunderstandingtheusecaseofmachinelearningintheautomotiveindustryan dhowmachinelearninghastransformedthedrivingexperience.WebuildaRandomForest Regressionmodeltopredicttheresalevalueofausedcar.Finally,weevaluatedtheperforma nceofthemodelusingtheRsquaredscoreandResidualPlot.
We could have also used simpler regression algorithms like Linear Regression and LassoRegression.Still,weneedtomakesuretherearenooutliersinthedatasetbeforeimple mentingthem.Pairplotsandscatterplotshelpvisualizetheoutliers.
ThenwehaveusedaFlaskapplicationtodisplaythepredictedvaluetotheusersbasedontheir corresponding input. This car resale value prediction can be used by the public to estimate theresalevalueofthecar.

12. FUTURESCOPE:

Currently,onlyfewfeaturesareusedtopredictresalevalueofthecar.Thiscanbeextended tomore features. One can also implement CNN to determine physical condition of the car
fromimageslikeidentifyingdents,scratchesetc.andthuspredictingmorerelevantresalevalu eofa
car.
For better performance, we plan to judiciously design deep learning network structures,
useadaptivelearningratesandtrainonclustersofdataratherthanthewholedataset.Tocorre ctforoverfittinginRandomForest,differentselectionsoffeaturesandnumberoftreeswillb etestedtocheckforchangeinperformance.

13. APPENDIX:

Sourcecode

app.py

```
importpandasaspdimportnu
mpyasnp
fromflaskimportFlask,render_template,Response,requestimportpickle
fromsklearn.preprocessingimportLabelEncoderimportpickle

app=Flask(name)
filename='resale_model(1).sav'
model_rand=pickle.load(open(filename,'rb'))


@app.route('/')defindex():
        returnrender_template('index.html')


@app.route('/home')defhom
e():
        returnrender_template('index.html')


@app.route('/predict')defpredict():
        returnrender_template('booking.html')


@app.route('/y_predict',methods=['GET','POST'])defy_predict():
        regyear =
        int(request.form['regyear'])powerps=float(request.form['power
        ps'])kms=float(request.form['kms'])
        regmonth=int(request.form.get('regmonth'))gearbox=request.form['gear
        box']
        damage=request.form['damaged']model=request.form.ge
        t('model_type')brand =
        request.form.get('brand')fuelType=request.form.get('fue
        l')
        vehicletype=request.form.get('vehicletype')
        new_row={'yearOfRegistration':regyear,'powerPS':powerps,'kilometer':kms,'monthOfRegistration':regmonth,
                        'gearbox':gearbox,'notRepairedDamage':damage,'model':model,'brand':brand,'fuelType':fuelType,
                        'vehicleType':vehicletype}
```

```
    print(new_row)
    new_df=pd.DataFrame(
        columns=['vehicleType','yearOfRegistration','gearbox','powerPS','model','kilometer','m
onthOfRegistration',
                 'fuelType','brand','notRepairedDamage'])new_df=n
    ew_df.append(new_row,ignore_index=True)
    labels=['gearbox','notRepairedDamage','model','brand','fuelType','vehicleType']mapper={}
    foriinlabels:
        mapper[i]=LabelEncoder()
        mapper[i].classes_=np.load(str('classes'+i+'.npy'),allow_pickle=True)tr=mapper
        [i].fit_transform(new_df[i])
        new_df.loc[:,i+'_Labels']=pd.Series(tr,index=new_df.index)labeled=new_d
    f[
        ['yearOfRegistration','powerPS','kilometer','monthOfRegistration']+[x+"_Labels"for
xinlabels]]

    X=labeled.valuesp
    rint(X)
    y_prediction=model_rand.predict(X)print(y_prediction)
    returnrender_template('booking.html',
                          ypred="Theresalevaluepredictedis₹{:.2f}
".format(y_prediction[0]))


ifname_____=='main':
    app.run(host='localhost',debug=True,threaded=False)
```

index.html

```
<!DOCTYPEhtml>
<html>
    <head>
        <style>



            }
        </style>
    </head>
    <body>
        <h1>CarResaleValuePrediction</h1>
```

```
        <h2>welcome</h2>
        <formaction='predict'>
            <p>Clickheretopredicttheprice</p>
            <inputtype="submit"value="Clickhere">
         </form>

    </body>
</html>
```

booking.html

```
<html>

   <head>
      <linkrel="stylesheet"href="styles.css">

      <linkrel="stylesheet"type="text/css"href="{{url_for('static',filename='styles/formstyle.c
ss')}}">
      <link rel="stylesheet" type="text/css"
href="{{url_for('static',filename='styles/font-
awesome.css')}}">


      <style>
            a{
                color:white;
            }
         .button{
            border:
            none;color:
            white;padding:15px
            32px;text-
            align:center;margin
            -top:20px;
            text-
            decoration:none;displa
            y:inline-block;font-
            size: 16px;margin:
            4px 2px;margin-top:
            2em;cursor:pointer;
         }

         body{
            padding:1em;
            font-family:"OpenSans","HelveticaNeue",Helvetica,Arial,sans-serif;font-
```

```
content/uploads/2014/06/supercar-wallpapers-bugatti-4.jpg');color:#b9b9b9;
                    background-color:black;background-repeat:no-repeat;background-size:cover;
                        background-attachment:fixed;
                }

                /*Green*/
                .button2{
                    background-color:white;color:black;
                        border-color:blacksolid;

                }
            </style>

        </head>

            <sectionclass="banner_main">
                <divclass="container">
                        <formaction="/y_predict"method="post">
                    <center>
                                        <h1style="color:beige;font-family:monospace";>PredictthePrice!</h1>
                    </center>

                    <center>
                        <h3style="font-family:verdana;color:white;">{{ypred}}</h3>
                    </center>
                    <divclass="row">
                        <h4>RegistrationDetails</h4>
                        <divclass="input-group"><inputtype="number"name="regyear"id="regyear"placeholder="RegistrationYear"required/>
                        </div>
                        <div>
                            <h4>RegistrationMonth</h4>
                            <!--<divclass="input-group"><inputtype="number"name="regmonth"
id="regmonth"
                                    placeholder="RegistrationMonth"/>-->

                                <divclass="input-group">
                                    <selectname="regmonth"id="regmonth">
                                        <optionvalue=1>January</option>
                                <optionvalue=2>February</option>
                                <optionvalue=3>March</option>
                                <optionvalue=4>April</option>
                                <optionvalue=5>May</option>
                                <optionvalue=6>June</option>
```

```
                                    <optionvalue=7>July</option>
                                    <optionvalue=8>August</option>
                                    <optionvalue=9>September</option>
                                    <optionvalue=10>October</option>
                                    <optionvalue=11>November</option>
                                    <optionvalue=12>Decemeber</option>
                                </select>
                            </div>

                            </div>

                    </div>
                        <h4>PowerPS</h4>
                        <divclass="input-group"><inputtype="number"name="powerps"placeholder="PoweroftheCarinPS"required/>
                        </div>
                        <h4>KilometersDriven</h4>
                        <divclass="input-group"><inputtype="number"name="kms"placeholder="Kilometersthecarhasdriven"required/>
                        </div>

                        <div>
                            <h4>GearBoxType</h4>
                            <divclass="input-group">
                                <inputid="gear-manual"type="radio"name="gearbox"value="manual"/>
                                <labelfor="gear-manual">Manual</label>
                                <inputid="gear-automatic"type="radio"name="gearbox"
value="automatic"/>
                                <labelfor="gear-automatic">Automatic</label>
                                <inputid="gear-notdeclared"type="radio"name="gearbox"value="not
declared"/>
                                <labelfor="gear-notdeclared">Notdeclared</label>

                            </div>
                        </div>

                        <div>
                            <h4>IsCarDamaged</h4>
                            <divclass="input-group">
                                <inputid="yes"type="radio"name="damaged"value="Yes"/>
                                <labelfor="yes">Yes</label>
                                <inputid="damaged-no"type="radio"name="damaged"value="No"/>
                                <labelfor="damaged-no">No</label>
                                <inputid="damaged-notdeclared"type="radio"name="damaged"value="Not
declared"/>
                                <labelfor="damaged-notdeclared">Notdeclared</label>
```

```
        </div>
    </div>

    <div>
        <h4>FuelType</h4>
        <divclass="input-group">
            <selectname="fuel"id="fuel">
                <optionvalue="not-declared">not-declared</option>
                <optionvalue="diesel">diesel</option>
                <optionvalue="petrol">petrol</option>
                <optionvalue="lpg">lpg</option>
                <optionvalue="others">others</option>
                <optionvalue="hybrid">hybrid</option>
                <optionvalue="cng">cng</option>
                <optionvalue="electric">electric</option>
            </select>
        </div>
    </div>


    <div>
        <h4>BrandoftheCar</h4>
        <divclass="input-group">
            <selectname="brand"id="brand">
                <optionvalue="audi">audi</option>
                <optionvalue="jeep">jeep</option>
                <optionvalue="Mahindra">Mahindra</option>
                <optionvalue="volkswagen">volkswagen</option>
                <optionvalue="skoda">skoda</option>
                <optionvalue="bmw">bmw</option>
                <optionvalue="peugeot">peugeot</option>
                <optionvalue="ford">ford</option>
                <optionvalue="mazda">mazda</option>
                <optionvalue="nissan">nissan</option>
                <optionvalue="renault">renault</option>
                <optionvalue="mercedes_benz">mercedes_benz</option>
                <optionvalue="honda">honda</option>
                <optionvalue="fiat">fiat</option>
                <optionvalue="opel">opel</option>
                <optionvalue="mini">mini</option>
                <optionvalue="smart">smart</option>
                <optionvalue="hyundai">hyundai</option>
                <optionvalue="alfa_romeo">alfa_romeo</option>
                <optionvalue="subaru">subaru</option>
                <optionvalue="volvo">volvo</option>
                <optionvalue="mitsubishi">mitsubishi</option>
                <optionvalue="kia">kia</option>
```

```
                    <optionvalue="seat">seat</option>
                    <optionvalue="lancia">lancia</option>
                    <optionvalue="porsche">porsche</option>
                    <optionvalue="citroen">citroen</option>
                    <optionvalue="toyota">toyota</option>
                    <optionvalue="chevrolet">chevrolet</option>
                    <optionvalue="dacia">dacia</option>
                    <optionvalue="suzuki">suzuki</option>
                    <optionvalue="daihatsu">daihatsu</option>
                    <optionvalue="chrysler">chrysler</option>
                    <optionvalue="sonstige_autos">sonstige_autos</option>
                    <optionvalue="jaguar">jaguar</option>
                    <optionvalue="daewoo">daewoo</option>
                    <optionvalue="rover">rover</option>
                    <optionvalue="saab">saab</option>
                    <optionvalue="land_rover">land_rover</option>
                    <optionvalue="lada">lada</option>
                    <optionvalue="trabant">trabant</option>

                </select>
            </div>
        </div>


        <div>
            <h4>ModelType</h4>
            <divclass="input-group">
                <selectname="model_type"id="model_type">
                    <optionvalue="not-declared">not-declared</option>
                    <optionvalue="grand">grand</option>
                    <optionvalue="scorpios11">scorpios11</option>
                    <optionvalue="golf">golf</option>
                    <optionvalue="fabia">fabia</option>
                    <optionvalue="3er">3er</option>
                    <optionvalue="2_reihe">2_reihe</option>
                    <optionvalue="c_max">c_max</option>
                    <optionvalue="3_reihe">3_reihe</option>
                    <optionvalue="passat">passat</option>
                    <optionvalue="navara">navara</option>
                    <optionvalue="polo">polo</option>
                    <optionvalue="twingo">twingo</option>
                    <optionvalue="a_klasse">a_klasse</option>
                    <optionvalue="scirocco">scirocco</option>
                    <optionvalue="5er">5er</option>
                    <optionvalue="andere">andere</option>
                    <optionvalue="civic">civic</option>
                    <optionvalue="punto">punto</option>
```

```
<optionvalue="e_klasse">e_klasse</option>
<optionvalue="clio">clio</option>
<optionvalue="kadett">kadett</option>
<optionvalue="one">one</option>
<optionvalue="fortwo">fortwo</option>
<optionvalue="1er">1er</option>
<optionvalue="b_klasse">b_klasse</option>
<optionvalue="a8">a8</option>
<optionvalue="jetta">jetta</option>
<optionvalue="c_klasse">c_klasse</option>
<optionvalue="micra">micra</option>
<optionvalue="vito">vito</option>
<optionvalue="sprinter">sprinter</option>
<optionvalue="astra">astra</option>
<optionvalue="156">156</option>
<optionvalue="escort">escort</option>
<optionvalue="forester">forester</option>
<optionvalue="xc_reihe">xc_reihe</option>
<optionvalue="fiesta">fiesta</option>
<optionvalue="scenic">scenic</option>
<optionvalue="ka">ka</option>
<optionvalue="a1">a1</option>
<optionvalue="transporter">transporter</option>
<optionvalue="focus">focus</option>
<optionvalue="a4">a4</option>
<optionvalue="tt">tt</option>
<optionvalue="a6">a6</option>
<optionvalue="jazz">jazz</option>
<optionvalue="omega">omega</option>
<optionvalue="slk">slk</option>
<optionvalue="7er">7er</option>
<optionvalue="combo">combo</option>
<optionvalue="corsa">corsa</option>
<optionvalue="80">80</option>
<optionvalue="147">147</option>
<optionvalue="glk">glk</option>
<optionvalue="z_reihe">z_reihe</option>
<optionvalue="sorento">sorento</option>
<optionvalue="ibiza">ibiza</option>
<optionvalue="mustang">mustang</option>
<optionvalue="eos">eos</option>
<optionvalue="touran">touran</option>
<optionvalue="getz">getz</option>
<optionvalue="insignia">insignia</option>
<optionvalue="almera">almera</option>
<optionvalue="megane">megane</option>
<optionvalue="a3">a3</option>
```

```
<optionvalue="r19">r19</option>
<optionvalue="caddy">caddy</option>
<optionvalue="mondeo">mondeo</option>
<optionvalue="cordoba">cordoba</option>
<optionvalue="colt">colt</option>
<optionvalue="impreza">impreza</option>
<optionvalue="vectra">vectra</option>
<optionvalue="lupo">lupo</option>
<optionvalue="berlingo">berlingo</option>
<optionvalue="m_klasse">m_klasse</option>
<optionvalue="tiguan">tiguan</option>
<optionvalue="6_reihe">6_reihe</option>
<optionvalue="c4">c4</option>
<optionvalue="panda">panda</option>
<optionvalue="up">up</option>
<optionvalue="i_reihe">i_reihe</option>
<optionvalue="ceed">ceed</option>
<optionvalue="kangoo">kangoo</option>
<optionvalue="5_reihe">5_reihe</option>
<optionvalue="yeti">yeti</option>
<optionvalue="octavia">octavia</option>
<optionvalue="zafira">zafira</option>
<optionvalue="mii">mii</option>
<optionvalue="rx_reihe">rx_reihe</option>
<optionvalue="6er">6er</option>
<optionvalue="modus">modus</option>
<optionvalue="fox">fox</option>
<optionvalue="matiz">matiz</option>
<optionvalue="beetle">beetle</option>
<optionvalue="rio">rio</option>
<optionvalue="touareg">touareg</option>
<optionvalue="logan">logan</option>
<optionvalue="spider">spider</option>
<optionvalue="cuore">cuore</option>
<optionvalue="s_max">s_max</option>
<optionvalue="a2">a2</option>
<optionvalue="x_reihe">x_reihe</option>
<optionvalue="a5">a5</option>
<optionvalue="galaxy">galaxy</option>
<optionvalue="c3">c3</option>
<optionvalue="viano">viano</option>
<optionvalue="s_klasse">s_klasse</option>
<optionvalue="1_reihe">1_reihe</option>
<optionvalue="sharan">sharan</option>
<optionvalue="avensis">avensis</option>
<optionvalue="sl">sl</option>
<optionvalue="roomster">roomster</option>
```

```
<optionvalue="q5">q5</option>
<optionvalue="santa">santa</option>
<optionvalue="leon">leon</option>
<optionvalue="cooper">cooper</option>
<optionvalue="4_reihe">4_reihe</option>
<optionvalue="sportage">sportage</option>
<optionvalue="laguna">laguna</option>
<optionvalue="ptcruiser">ptcruiser</option>
<optionvalue="clk">clk</option>
<optionvalue="primera">primera</option>
<optionvalue="espace">espace</option>
<optionvalue="exeo">exeo</option>
<optionvalue="159">159</option>
<optionvalue="transit">transit</option>
<optionvalue="juke">juke</option>
<optionvalue="v40">v40</option>
<optionvalue="carisma">carisma</option>
<optionvalue="accord">accord</option>
<optionvalue="corolla">corolla</option>
<optionvalue="lanos">lanos</option>
<optionvalue="phaeton">phaeton</option>
<optionvalue="boxster">boxster</option>
<optionvalue="verso">verso</option>
<optionvalue="rav">rav</option>
<optionvalue="kuga">kuga</option>
<optionvalue="qashqai">qashqai</option>
<optionvalue="swift">swift</option>
<optionvalue="picanto">picanto</option>
<optionvalue="superb">superb</option>
<optionvalue="stilo">stilo</option>
<optionvalue="alhambra">alhambra</option>
<optionvalue="911">911</option>
<optionvalue="m_reihe">m_reihe</option>
<optionvalue="roadster">roadster</option>
<optionvalue="ypsilon">ypsilon</option>
<optionvalue="galant">galant</option>
<optionvalue="justy">justy</option>
<optionvalue="90">90</option>
<optionvalue="sirion">sirion</option>
<optionvalue="signum">signum</option>
<optionvalue="crossfire">crossfire</option>
<optionvalue="agila">agila</option>
<optionvalue="duster">duster</option>
<optionvalue="v50">v50</option>
<optionvalue="mx_reihe">mx_reihe</option>
<optionvalue="meriva">meriva</option>
<optionvalue="discovery">discovery</option>
```

```
<optionvalue="c_reihe">c_reihe</option>
<optionvalue="v_klasse">v_klasse</option>
<optionvalue="yaris">yaris</option>
<optionvalue="c5">c5</option>
<optionvalue="aygo">aygo</option>
<optionvalue="seicento">seicento</option>
<optionvalue="cc">cc</option>
<optionvalue="carnival">carnival</option>
<optionvalue="fusion">fusion</option>
<optionvalue="bora">bora</option>
<optionvalue="cl">cl</option>
<optionvalue="tigra">tigra</option>
<optionvalue="300c">300c</option>
<optionvalue="500">500</option>
<optionvalue="100">100</option>
<optionvalue="q3">q3</option>
<optionvalue="cr_reihe">cr_reihe</option>
<optionvalue="spark">spark</option>
<optionvalue="x_type">x_type</option>
<optionvalue="ducato">ducato</option>
<optionvalue="s_type">s_type</option>
<optionvalue="x_trail">x_trail</option>
<optionvalue="toledo">toledo</option>
<optionvalue="altea">altea</option>
<optionvalue="voyager">voyager</option>
<optionvalue="calibra">calibra</option>
<optionvalue="v70">v70</option>
<optionvalue="bravo">bravo</option>
<optionvalue="range_rover">range_rover</option>
<optionvalue="forfour">forfour</option>
<optionvalue="tucson">tucson</option>
<optionvalue="q7">q7</option>
<optionvalue="c1">c1</option>
<optionvalue="citigo">citigo</option>
<optionvalue="jimny">jimny</option>
<optionvalue="cx_reihe">cx_reihe</option>
<optionvalue="cayenne">cayenne</option>
<optionvalue="wrangler">wrangler</option>
<optionvalue="lybra">lybra</option>
<optionvalue="range_rover_sport">range_rover_sport</option>
<optionvalue="lancer">lancer</option>
<optionvalue="freelander">freelander</option>
<optionvalue="captiva">captiva</option>
<optionvalue="range_rover_evoque">range_rover_evoque</option>
<optionvalue="sandero">sandero</option>
<optionvalue="note">note</option>
<optionvalue="antara">antara</option>
```

```
                        <optionvalue="900">900</option>
                        <optionvalue="defender">defender</option>
                        <optionvalue="cherokee">cherokee</option>
                        <optionvalue="clubman">clubman</option>
                        <optionvalue="arosa">arosa</option>
                        <optionvalue="legacy">legacy</option>
                        <optionvalue="pajero">pajero</option>
                        <optionvalue="auris">auris</option>
                        <optionvalue="c2">c2</option>
                        <optionvalue="niva">niva</option>
                        <optionvalue="s60">s60</option>
                        <optionvalue="nubira">nubira</option>
                        <optionvalue="vivaro">vivaro</option>
                        <optionvalue="g_klasse">g_klasse</option>
                        <optionvalue="lodgy">lodgy</option>
                        <optionvalue="850">850</option>
                        <optionvalue="serie_2">serie_2</option>
                        <optionvalue="charade">charade</option>
                        <optionvalue="croma">croma</option>
                        <optionvalue="outlander">outlander</option>
                        <optionvalue="gl">gl</option>
                        <optionvalue="kaefer">kaefer</option>
                        <optionvalue="doblo">doblo</option>
                        <optionvalue="musa">musa</option>
                        <optionvalue="amarok">amarok</option>
                        <optionvalue="9000">9000</option>
                        <optionvalue="kalos">kalos</option>
                        <optionvalue="v60">v60</option>
                        <optionvalue="200">200</option>
                        <optionvalue="145">145</option>
                        <optionvalue="b_max">b_max</option>
                        <optionvalue="delta">delta</option>
                        <optionvalue="aveo">aveo</option>
                        <optionvalue="rangerover">rangerover</option>
                        <optionvalue="move">move</option>
                        <optionvalue="materia">materia</option>
                        <optionvalue="terios">terios</option>
                        <optionvalue="kalina">kalina</option>
                        <optionvalue="elefantino">elefantino</option>
                        <optionvalue="i3">i3</option>
                        <optionvalue="samara">samara</option>
                        <optionvalue="kappa">kappa</option>
                        <optionvalue="serie_3">serie_3</option>
                        <optionvalue="discovery_sport">discovery_sport</option>
                    </select>
                </div>
            </div>
```

```
            <div>
               <h4>VehicleType</h4>
               <divclass="custom">
                     <selectname="vehicletype"id="wgtmsr">
                       <optionvalue="coupe">coupe</option>
                     <optionvalue="suv">suv</option>
                     <optionvalue="smallcar">smallcar</option>
                     <optionvalue="limousine">limousine</option>
                     <optionvalue="convertible">convertible</option>
                     <optionvalue="bus">bus</option>
                     <optionvalue="combination">combination</option>
                     <optionvalue="not-declared">not-declared</option>
                     <optionvalue="others">others</option>
                  </select>
               </div>
            </div>


            <divclass="row">
               <center><buttonclass="buttonbutton2"href="/"type="submit">Predict</button></
center>
            </div>

            <br>
            <center><ahref="/home">ClickheretogobackHome!</a></center>
         </form>

      </div>
   </section>
   </html>
```

Githublink: IBM-EPBL/IBM-Project-42049-1660648019

Projectdemolink

https://github.com/IBM-EPBL/IBM-Project-42049-1660648019/tree/main/Final%20Deliverables