CAR RESALE VALUE PREDICTION

IBM-Project-26582-1660029937

Team ID: PNT2022TMID50997

A PROJECT REPORT

Submitted by

ASWINKUMAR M	**(953619104005)**
ABHINAV R	**(953619104001)**
THARUNKUMAR B	**(953619104044)**
SRIRAMAKRISHNAN J	**(953619104040)**

CHAPTER 1
INTRODUCTION

Given the variety of elements that influence a used car's market pricing, determining if the quoted price is accurate is a difficult undertaking. The goal of this research is to create machine learning models that can precisely forecast a used car's price based on its attributes so that buyers can make educated decisions. Implement and assess various learning techniques using a dataset of sale prices for various brands and models. The cost of the car will be determined based on a number of factors. Regression It is feasible to forecast the actual price of a car rather than just the price by using algorithms because they provide us a continuous value rather than a classified value as an output.

1.1 Project Overview

- Able to understand the problem to classify if it is a regression or a classification kind of problem.
- Able to know how to pre-process/clean the data using different data preprocessing techniques.
- Applying different algorithms according to the dataset.
- Able to know how to evaluate the model.
- Able to build web applications using the Flask framework.

1.2 Purpose

Car resale value prediction system is made with the purpose of predicting the correct valuation of used cars that helps users to sell the car remotely with perfect valuation and without human intervention in the process to eliminate biased valuation. Used car resale market in India was marked at 24.2 billion US dollars in 2019. Due to the huge requirement of used cars and lack of experts who can determine the correct valuation, there is an utmost need of bridging this gap between sellers and buyers. This project focuses on building a system that can accurately predict a resale value of the car based on minimal features like kms driven, year of purchase etc. without manual or human interference and hence it remains unbiased.

CHAPTER 2 LITERATURE SURVEY

2.1 Existing problem

Using various machine learning algorithms we will predict the price. The algorithms involves Random Forest Regression. The best model which predicts the most accurate price is selected. After selection of the best model the predicted price is displayed to the user according to user's inputs. User can give input through website to for used car price prediction to machine learning model.

2.2 References

Authors Title and Publication	Technology	Advantage	Disadvantage
Praful Rane,Deep Pandya,Dhawal Kotak, 'USED CAR PRICE PREDICTION', International Research Journal of Engineering and Technology (IRJET) Published in April 4,2021	Random Forest Regression, Hyper parameters Random forest, Extra-tree Regressor	It is estimated that a car's value depriciated to almost 50%. The calculation for the first year is about 20%. That means if you buy a used car that was launched a year ago, you can save upto 20% on its original cost.	Even we curated a list of car that are available 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 in huge repair costs while you use them.
Enis Gegic,Becir Isakovic,Dino Keco Zerina Masetic, Jasmin Kevric, 'Car Price Prediction using Machine Learning Techniques', TEM Journal Published in 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 a 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.

Mukkesh Ganesh, Pattabiraman 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 authorised dealership, you get a warranty on the repair. Read this carefully. If u buy a car from a True Value dealership and the car will be 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 ends too. Although the prices are obviously lower in the used car markets, you don't get that much "discounts".
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 lack of experts who can determine the correct valuation, there is an utmost need of bridging this gap between sellers and buyers.
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.
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 neighbours , naive bayes and decision trees	This is the major differential between a new car and a used car. For the same cost of 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 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.

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.

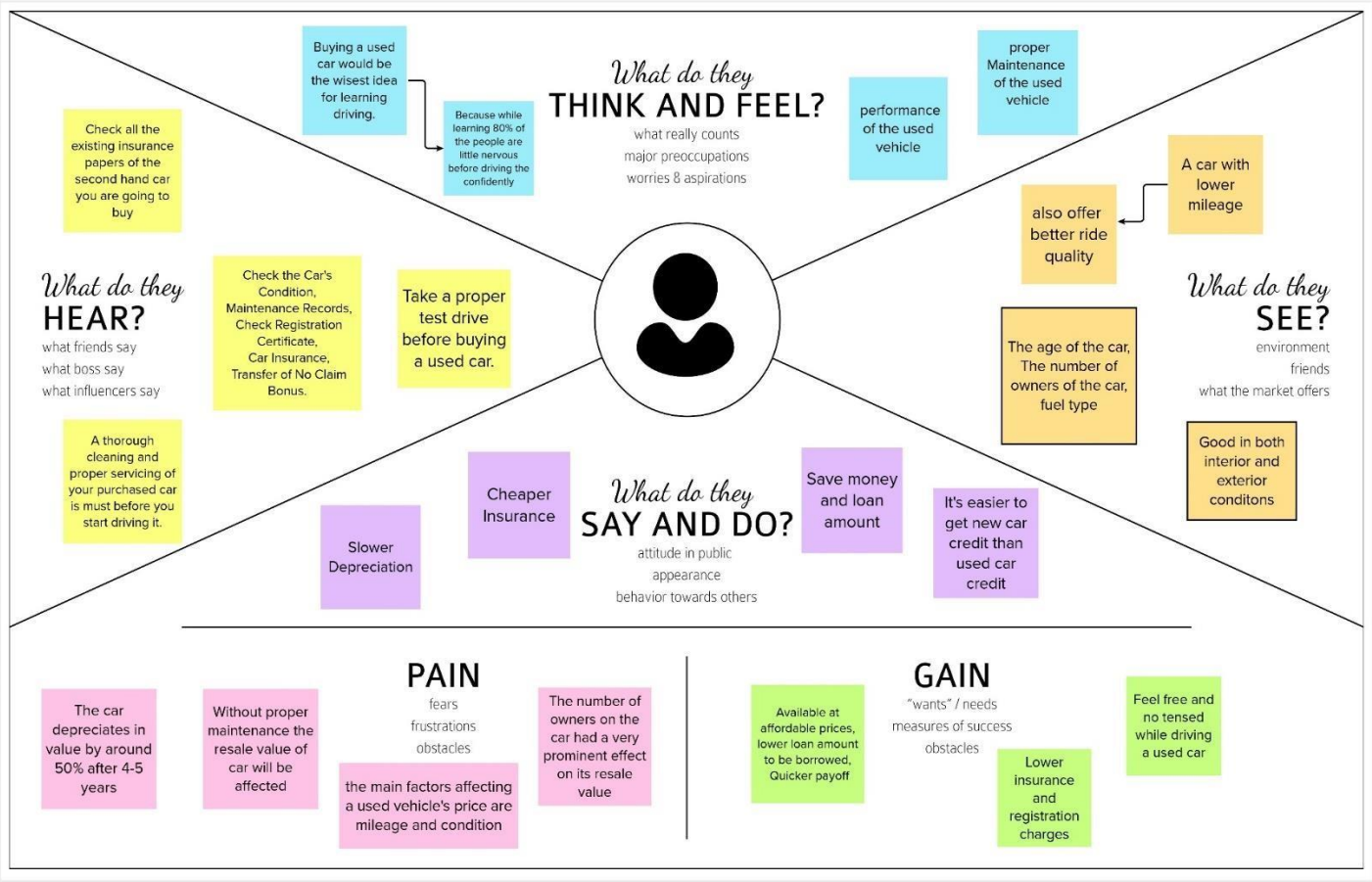
2.3 Problem Statement Definition

The prices of new cars in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So customers buying a new car can be assured of the money they invest to be worthy. But due to the increased price of new cars and the incapability of customers to buy new cars due to the lack of funds, used cars sales are on a global increase. Predicting the prices of used cars is an interesting and much-needed problem to be addressed. Customers can be widely exploited by fixing unrealistic prices for the used cars and many falls into this trap. Therefore, rises an absolute necessity of a used car price prediction system to effectively determine the worthiness of the car using a variety of features. Due to the adverse pricing of cars and the nomadic nature of people in developed countries, the cars are mostly bought on a lease basis, where there is an agreement between the buyer and seller. These cars upon completion of the agreement are resold. So reselling has become an essential part of today's world.

The main aim of this project is to predict the price of used cars using the various Machine Learning (ML) models. This can enable the customers to make decisions based on different inputs or factors namely Brand or Type of the car one prefers like Ford, Hyundai, Model of the car namely Ford Figo, Hyundai Creta, Year of manufacturing like 2020, 2021, Type of fuel namely Petrol, Diesel, Price range or Budget, Type of transmission which the customer prefers like Automatic or Manual, Mileage to name a few characteristic features required by the customer. This project Car Price Prediction deals with providing the solution to these problems. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. The predictions are then evaluated and compared in order to find those which provide the best performances.

CHAPTER 3 IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation and Brainstorming

MEMBER 1:

- Car Prediction using image. ○ By using exterior and interior image of the car.
- The value will be predicted based on the appearance of the car.If there any damage or n numbers scratches the car resale value will be quite affected. ○ By using neural network value of the car can be predicted
- Neural network algorithm is developed by considering the human brain that takes a set of unit as input and transfers results to a predefined output

MEMBER2:

- The main objective of this project is to *predict* the *Prices* of *used cars*, compare the ***prices*** and also estimate the *life span* of a particular ***car.***
- Insurance, Company claims,etc

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

MEMBER 3: ○ Car prediction using engine condition. ○ user should Upload engine Sound in the format of audio file. ○ By using Convolutional Neural Networks methodology price can be predicted.

- CNNs for Machine Learning on sound data by spectrogram approach that was just converts each song (or song segment) into a spectrogram: a twodimensional matrix

MEMBER 4:

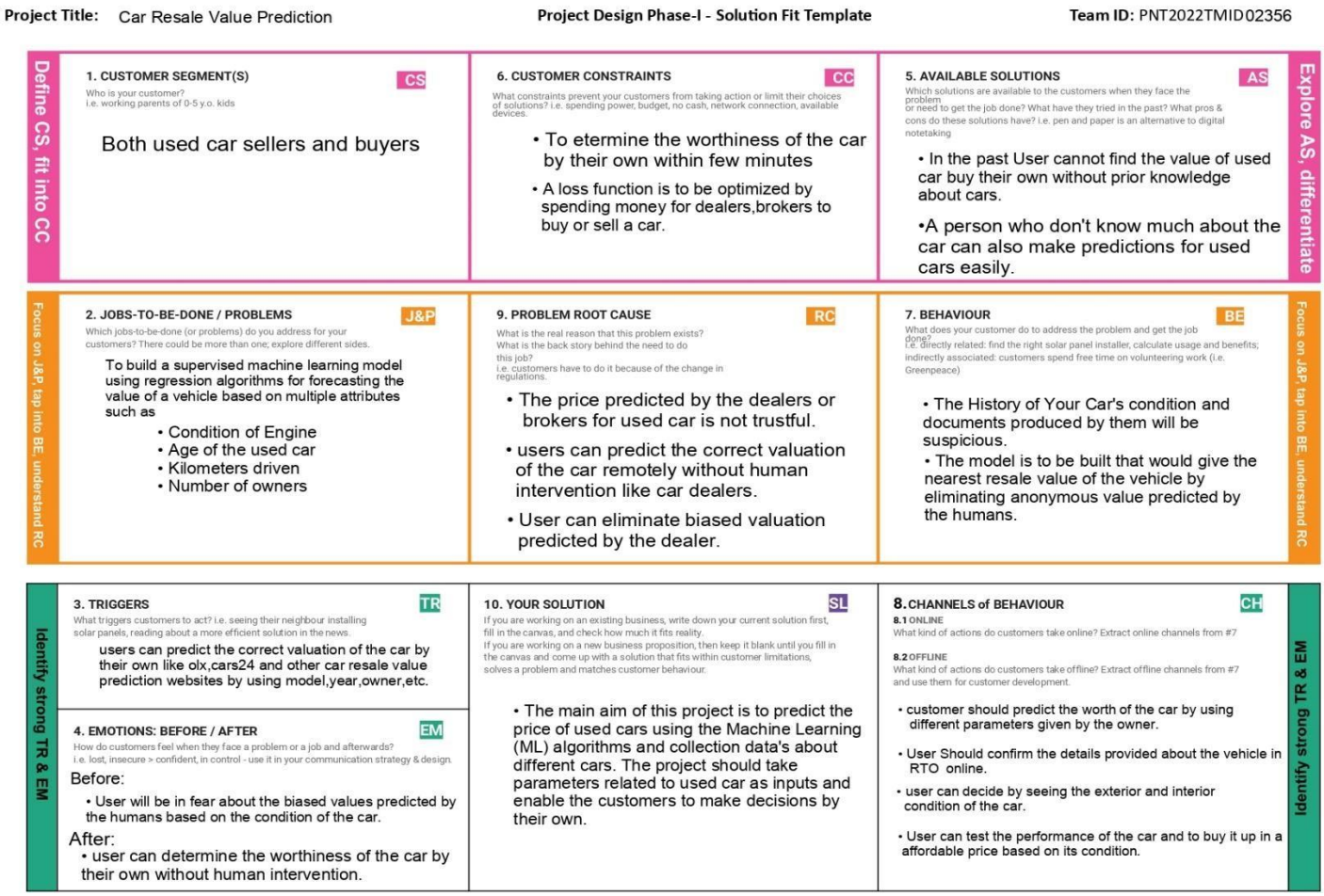
- Economic Conditions. ○ Kilometres Covered.
- Its mileage (the number of kilometers it has run) and its horsepower ○ Car prediction using XGBoost algorithm accurate result will be monitored.
- XGBoost as a regression model gave the best MSLE and RMSLE values.

3.3 Proposed Solution

S. No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The main aim of this project is to predict the price of used cars using the various Machine Learning (ML) models. The project should take parameters related to used car as inputs and enable the customers to make decisions by their own.
2.	Idea / Solution description	The model is to be built that would give the nearest resale value of the vehicle. By using these best accuracy value will be taken as a solution and it will be integrated to the webbased application where the user is notified with the status of his product.
3.	Novelty / Uniqueness	Used car price prediction is effectively used to determine the worthiness of the car by their own within few minutes by using various features such as year, model, mileage(km), etc.

4.	Social Impact / Customer Satisfaction	• If the user wants to buy or sell a own car it helps users to predict the correct valuation by their own. • A loss function is to be optimized and mainly a weak learner can make predictions for used cars easily.
5.	Business Model (Revenue Model)	• It helps users to predict the correct valuation of the car remotely with perfect valuation and without human intervention like car dealers in the process to eliminate biased valuation predicted by the dealer.
6.	Scalability of the Solution	• Using Stored data and machine learning approaches, this project proposed a scalable framework for predicting values for different type of used cars present all over India.

3.4 Problem Solution fit



CHAPTER 4
REQUIREMENT ANALYSIS

4.1 Functional requirement

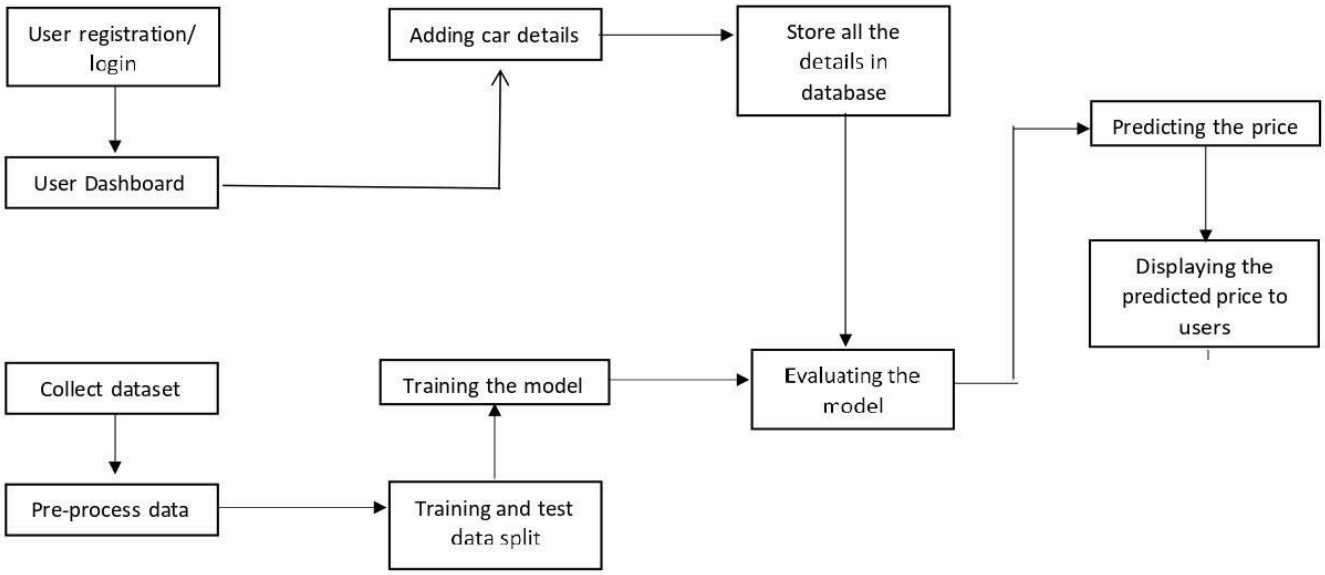
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Gmail.
FR-2	User Confirmation	Confirmation via Email.
FR-3	User Login Window	Login using given credentials.
FR-4	Dashboard	Fill the required data shown in the window.
FR-5	Prediction Value	Predicting the value of the car and displayed in the window.

4.2 Non-Functional requirements

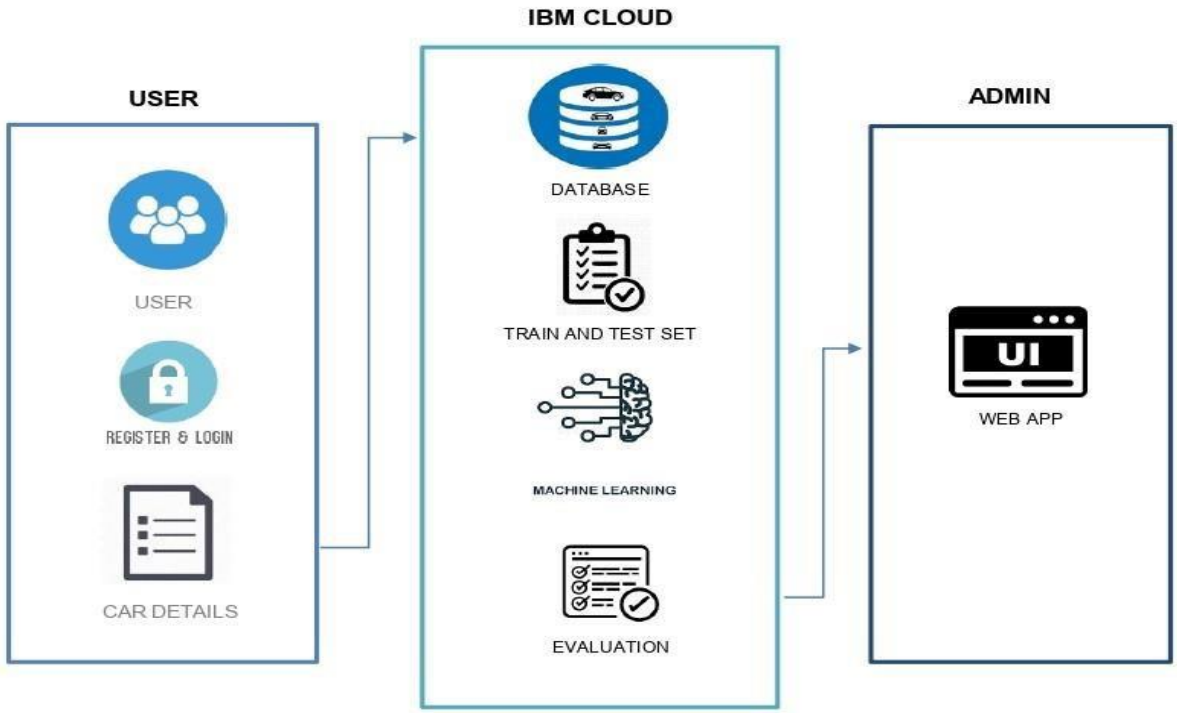
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Predicting the value of used cars.
NFR-2	**Security**	Aware about fraudulent sites the data's given by the user is not exposed in any way.
NFR-3	**Reliability**	It helps user to predict the correct valuation of the car remotely with perfect valuation and without human intervention like car dealers.
NFR-4	**Performance**	Users can determine the worthiness of the car by their own within a few minutes.
NFR-5	**Availability**	It is available for everyone and can be accessed anywhere at anytime.
NFR-6	**Scalability**	This project proposed a scalable framework for predicting values for different types of used cars present all over India.

CHAPTER 5 PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution and Technical Architecture



5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Admin	Dataset	USN-1	Gather the information needed to make the car resale prediction.	enough data has beengathered to train the model.	High	Sprint-1
	Data preprocessing	USN-2	Perform data cleaning to optimize the dataset	Clean Dataset enough tomake correct predictions	High	Sprint-1
	Training & Building Model	USN-3	Build the model using regression algorithmsto classify the data	Model should be used forPredicting perfect valuation of the car.	High	Sprint-1
	Deploy the model	USN-4	Deployment of ML model using IBM Cloud	Model should be workingfine from the cloud	High	Sprint-2
	Integrate the web app with the IBM model	USN-5	Use flask for the integration purpose.	The model ought to be simple to use and reliable onthe web application.	High	Sprint-2
Customer	Homepage	USN-6	Information on the application and the process for selling used cars	We can get an idea abouthow to use these model.	Medium	Sprint-2
	Registration	USN-7	User can register the application by entering email, password, and confirming my password.	user can access myaccount /dashboard	High	Sprint-3
		USN-8	user will receive confirmation email.jonce they have registered for the application	user can receive confirmationemail	High	Sprint-3
	Login	USN-9	user can log into the application byentering email & password	user can login to myaccount	High	Sprint-3
	Dashboard	USN-10	User can add new cars and get accessto insert and update their details	user can add new cars	Medium	Sprint-4
	Car Details	USN-11	user should give the required car details like carmodel, kilometer driven,manufactured year, etc...	After filling the car'sdetails and taking them for furtherprocessing.	High	Sprint-4
	Car Price	USN-12	The price of a used car can be displayed and seen by the user.	Car prices must be displayed depending on thedata provided by the user.	High	Sprint-4

CHAPTER 6
PROJECT PLANNING AND SCHEDULING

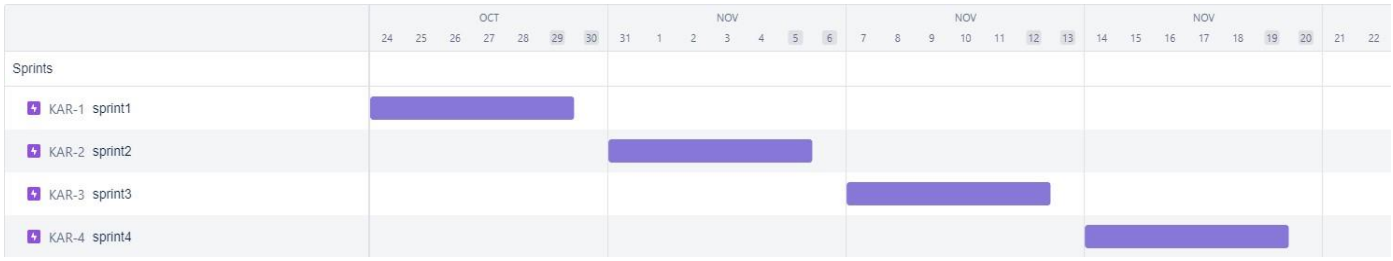
6.1 Sprint Planning and Estimation

User	Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Admin	Sprint 1	Dataset collection	USN-1	Collect the required data for the Car resale prediction	2	High
	Sprint 1	Data preprocessing	USN-2	Perform data cleaning to optimize the dataset	4	Medium
	Sprint 1	Training & Building Model	USN-3	Build the model using regression algorithms to classify the data	6	High
	Sprint 2	Deploy the model	USN-4	Deployment of ML model using IBM Cloud	5	High
	Sprint 4	Integration	USN-5	Integrate the web app developed using flask with IBM model	5	High
Customer	Sprint 2	Homepage	USN-6	Details about the application and the car resale process	2	Low
	Sprint 2	Registration	USN-7	As a user, I can register for the application by entering confirming.	5	High
	Sprint 3	Confirmation	USN-8	As a user, I will receive confirmation email once I have registered for the application	3	Medium
	Sprint 3	Login	USN-9	As a user, I can log into the application by entering email & password	4	High
	Sprint 3	Dashboard	USN-10	As a user, I can add new cars and get access to insert and update their details	5	High
	Sprint 4	Car Details	USN-11	As a user, I should give the car details like car model, engine and fuel type, etc…	2	Medium
	Sprint 4	Car Price	USN-12	As a user, I can view the current rate of the used car price	5	High

6.2 Sprint Delivery Schedule

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date
Sprint-1	20	6 Days	03 Oct 2022	09 Nov 2022	20	09 Nov 2022
Sprint-2	20	6 Days	04 Nov 2022	10 Nov 2022	20	10 Nov 2022
Sprint-3	20	6 Days	06 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	13 Nov 2022	19 Nov 2022	20	19 Nov 2022

6.3 Reports from JIRA



Roadmap

CHAPTER 7
CODING AND SOLUTIONING

7.1 Feature 1

Regression Model:

Random Forest Regressor

```
In [20]: from sklearn.ensemble import RandomForestRegressor
         from sklearn.metrics import r2_score
         regressor = RandomForestRegressor(n_estimators=1000,max_depth=10,random_state=34)
         regressor.fit(X_train,np.ravel(Y_train,order='C'))
         X_test

Out[20]: array([[  2008,    140, 150000, ...,      1,      1,      4],
                [  2000,     60, 150000, ...,     38,      7,      7],
                [  2004,     54, 125000, ...,     38,      7,      7],
                ...,
                [  2001,    102, 150000, ...,     20,      7,      4],
                [  2003,    146, 150000, ...,      2,      7,      2],
                [  1999,     75, 150000, ...,     24,      7,      7]], dtype=int64)

In [17]: y_pred=regressor.predict(X_test)
         y_pred

Out[17]: array([9937.11993609, 1157.33268937, 2739.06719477, ..., 2124.91777611,
                2955.66758818, 1207.44132337])

In [18]: print(r2_score(Y_test,y_pred))

         0.834527626497731

In [19]: filename='resale_model.sav'
         pickle.dump(regressor,open(filename,'wb'))
```

7.2 Feature 2 Accuracy:

```
In [40]: from sklearn.ensemble import RandomForestRegressor
         from sklearn.metrics import r2_score
         regressor = RandomForestRegressor(n_estimators=1000,max_depth=10,random_state=34)
         regressor.fit(X_train,np.ravel(Y_train,order='C'))

Out[40]: RandomForestRegressor(max_depth=10, n_estimators=1000, random_state=34)

In [41]: y_pred=regressor.predict(X_test)
         print(r2_score(Y_test,y_pred))

         0.834527626497731

In [42]: filename='resale_model.sav'
         pickle.dump(regressor,open(filename,'wb'))
```

CHAPTER 8
TESTING

8.1 Test Cases

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status
/elcome_TC_OO1	Functional	Welcome Page	On Successful redirecting to flask application it must appear before the main page.	Should have downloaded a browser.	1.Enter URL and click go 2.The Page with information about car resale and project appear.		Welcome page should popup should display	Working as expected	Pass
unctional_TC_OO 2	UI	Home Page	Verify the UI elements in main page	Should have downloaded a browser with internet connection.	1.Enter URL and click go 2.Verify below UI elements: a.check the car type dropdown. b.check the car fuel type dropdown. c.check the year dropdown. d.check the ownertype dropdown.		Application should show below UI elements: a.car type dropdown b.car fuel type dropdown. c.year dropdown. d.ownertype dropdown.	Working as expected	Pass
Predicted value_TC_OO3	Functional/Result	Notificatio n	Verify user is able to get the value for the given details.	Browser with network condition and no fault in execution.	Enter the details and ensure the following. 1.Select appropriate value in dropdown. 2.Enter numbers in required integer field. 3.Enter letters in the appropriate character Field.	password: Testing123	User should get the predicted value as notification below the page.	Giving the expected value	Pass

8.2 User Acceptance Testing

1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the [ProductName] project at the time of the release to User Acceptance Testing (UAT).

2. Defect Analysis

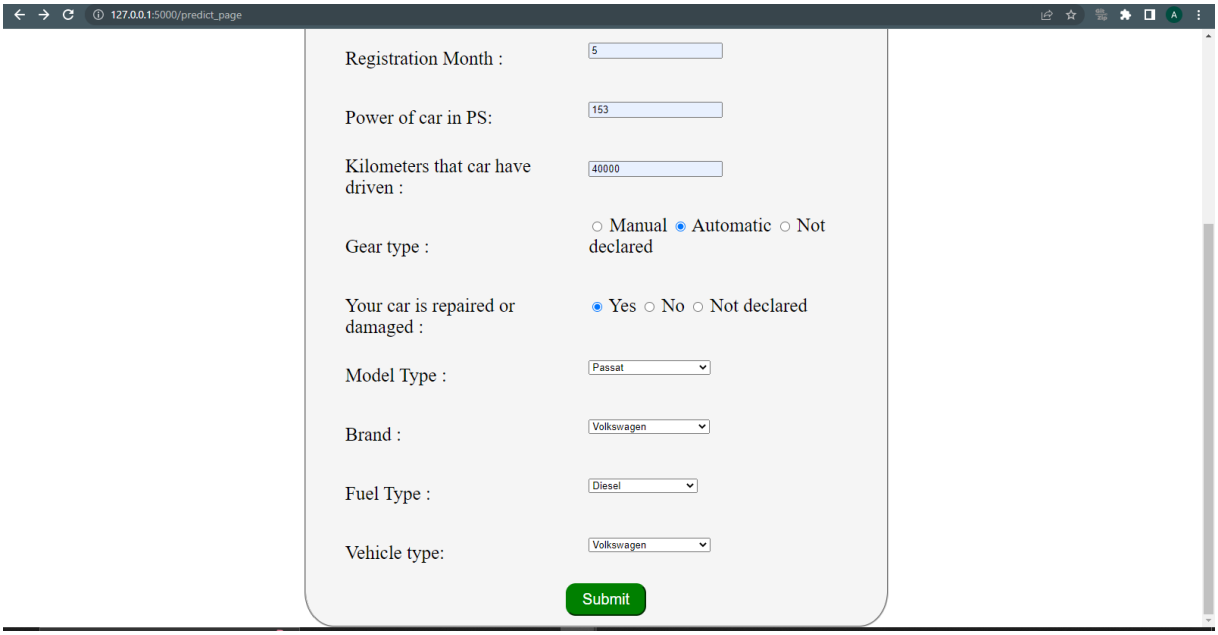
This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

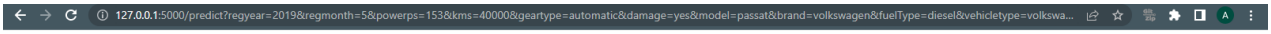
Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

3. Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3



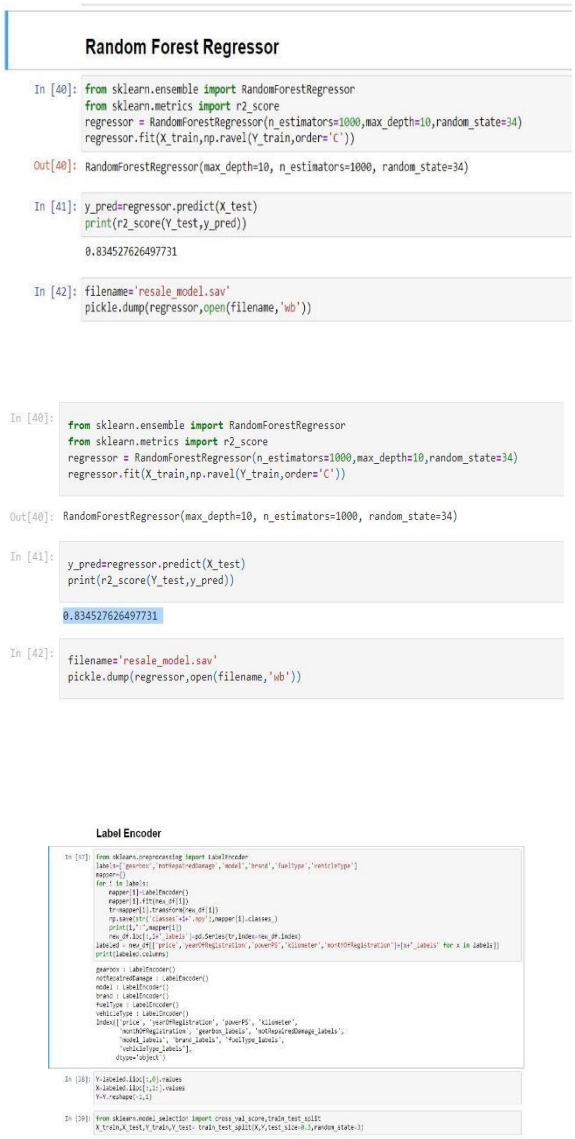


The Predicted Car Resale Value is

₹[662443.00724512]

CHAPTER 9
RESULTS

9.1 Performance Metrics

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Regression Model:** MAE - , MSE - , RMSE - , R2 score - **Classification Model:** Confusion Matrix - , Accuray Score- & Classification Report – Label Encoder	

CHAPTER 10

ADVANTAGES

- The model is to be built that would give the nearest resale value of the vehicle.
- By using these model best accuracy value will be taken as a solution for the given used car.
- Used car price prediction is effectively used to determine the worthiness of the car by their own within few minutes by giving different parameters as input such as year, model, km driven, etc.
- A loss function is to be optimized and mainly a weak learner can make predictions for used cars easily.
- Using Stored data and machine learning approaches, this project proposed a scalable framework for predicting values for different type of used cars present all over India.
- It helps users to predict the correct valuation of the car remotely with perfect valuation and without human intervention like car dealers in the process to eliminate biased valuation predicted by the dealer

DISADVANTAGES

- Even though predicting the value of a used car by these model Taking a proper test drive before buying a used car will show a perfect condition of the car to predict the value by their own.
- The main limitation of random forest is that a large number of trees can make the algorithm too slow and ineffective for real-time predictions.
- In general, these algorithms are fast to train, but quite slow to create predictions once they are trained.
- They are largely unstable compared to other decision predictors.

CHAPTER 11 CONCLUSION

The increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on a global increase. Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of used car price prediction.

CHAPTER 12 FUTURE SCOPE

In future this machine learning model may bind with various website 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 user interface for interacting with user. 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.

CHAPTER 13 APPENDIX

Source Code Model building:

```
import pandas as pd
import numpy as np import
matplotlib as plt
from sklearn.preprocessing import LabelEncoder
import pickle import os, types import pandas as
pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
cos_client = ibm_boto3.client(service_name='s3',
ibm_api_key_id='xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx',
ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
   config=Config(signature_version='oauth'),    endpoint_url='https://s3.private.us.cloud-object-
storage.appdomain.cloud')

bucket = 'carresalevaluepredictiondeploymen-donotdelete-pr-ryosh4pvhxemjh' object_key
= 'autos_preprocessed.csv'

body = cos_client.get_object(Bucket=bucket,Key=object_key)['Body'] #
add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

df = pd.read_csv(body) df.head()

# print(df.seller.value_counts())
# df[df.seller!='gewerblich']
# # df=df.drop('seller',1)
# print(df.offerType.value_counts())
# df[df.offerType!='Gesuch']
# df=df.drop('offerType',1)
```

```
print(df.shape)
df=df[(df.powerPS >50) & (df.powerPS <900)]
df=df[(df.yearOfRegistration >=1950) & (df.yearOfRegistration <2017)]
print(df.shape) (278578, 12)
(278578, 12)

# df.drop(['name','abtest','dateCrawled','nrOfPictures','lastSeen','postalCode','dateCreated'],axis='columns',inplace=True)
new_df=df.copy()
new_df=new_df.drop_duplicates(['price','vehicleType','yearOfRegistration','gearbox','powerPS','model','kilometer','monthOfRegistr
ation','fuelType','notRepairedDamage'])
 new_df.gearbox.replace(('manuell','automatik'),('manual','automatic'),inplace=True)
new_df.fuelType.replace(('benzin','andere','elektro'),('petrol','others','electric'),inplace=True)
new_df.vehicleType.replace(('kleinwagen','cabrio','kombi','andere'),('small car','convertible','combination','others'),inplace=True)
new_df.notRepairedDamage.replace(('ja','nein'),('Yes','No'),inplace=True)
new_df=new_df[(new_df.price >=100)&(new_df.price <=150000)]
new_df['notRepairedDamage'].fillna(value='not-declared',inplace=True)
new_df['fuelType'].fillna(value='not-declared',inplace=True)
new_df['gearbox'].fillna(value='not-declared',inplace=True) #
new_df['vehicleType'].fillna(value='not-declared',inplace=True)
new_df['model'].fillna(value='not-declared',inplace=True)

# new_df.to_csv("autos_preprocessed.csv")
 from sklearn.preprocessing import LabelEncoder
labels=['gearbox','notRepairedDamage','model','brand','fuelType','vehicleType']
mapper={} for i in labels:
mapper[i]=LabelEncoder()
mapper[i].fit(new_df[i])
tr=mapper[i].transform(new_df[i])
 np.save(str('classes'+i+'.npy'),mapper[i].classes_)
print(i,":",mapper[i])
new_df.loc[:,i+'_labels']=pd.Series(tr,index=new_df.index)
labeled = new_df[['price','yearOfRegistration','powerPS','kilometer','monthOfRegistration']+[x+"_labels" for x in labels]]
print(labeled.columns) gearbox : LabelEncoder() notRepairedDamage : LabelEncoder() model : LabelEncoder() brand :
LabelEncoder() fuelType : LabelEncoder() vehicleType : LabelEncoder()
Index(['price', 'yearOfRegistration', 'powerPS', 'kilometer',
       'monthOfRegistration', 'gearbox_labels', 'notRepairedDamage_labels',
       'model_labels', 'brand_labels', 'fuelType_labels',
       'vehicleType_labels'],
dtype='object')

Y=labeled.iloc[:,0].values
X=labeled.iloc[:,1:].values
Y=Y.reshape(-1,1)
from sklearn.model_selection import cross_val_score,train_test_split 18
X_train,X_test,Y_train,Y_test= train_test_split(X,Y,test_size=0.3,random_state=3)
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
```

```
regressor = RandomForestRegressor(n_estimators=1000,max_depth=10,random_state=34)
regressor.fit(X_train,np.ravel(Y_train,order='C'))

RandomForestRegressor(max_depth=10, n_estimators=1000, random_state=34)

y_pred=regressor.predict(X_test) print(r2_score(Y_test,y_pred))
0.834527626497731
filename='resale_model.sav'
pickle.dump(regressor,open(filename,'wb'))
```

Git Repo - https://github.com/IBM-EPBL/IBM-Project-26582-1660029937