PROJECT REPORT - 2022

CAR RESALE VALUE PREDICTION

IBM Team ID: PNT2022TMID20203

TEAM MEMBERS:

MARY SHIPANI J - 49621912078

PRIYADARSHINI G- 49621912090

SHENBAGALAKSHMI S - 49621912102

KALAISELVI P– 49621912122

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

NATIONAL ENGINEERING COLLEGE

(An Autonomous institution affiliated to Anna University)

KOVILPATTI - 628501

Project Report Format

1. **INTRODUCTION**
    1. Project Overview
    2. Purpose
2. **LITERATURE SURVEY**
    1. Existing problem
    2. References
    3. Problem Statement Definition
3. **IDEATION & PROPOSED SOLUTION**
    1. Empathy Map Canvas
    2. Ideation & Brainstorming
    3. Proposed Solution
    4. Problem Solution fit
4. **REQUIREMENT ANALYSIS**
    1. Functional requirement
    2. Non-Functional requirements
5. **PROJECT DESIGN**
    1. Data Flow Diagrams
    2. Solution & Technical Architecture
    3. User Stories
6. **PROJECT PLANNING & SCHEDULING**
    1. Sprint Planning & Estimation
    2. Sprint Delivery Schedule
    3. Reports from JIRA
7. **CODING & SOLUTIONING (Explain the features added in the project along with code)**
    1. Feature 1
    2. Feature 2
    3. Database Schema (if Applicable)
8. **TESTING**
    1. Test Cases
    2. User Acceptance Testing
9. **RESULTS**
    1. Performance Metrics
10. **ADVANTAGES & DISADVANTAGES**
11. **CONCLUSION**
12. **FUTURE SCOPE**
13. **APPENDIX**

    13.1 Source Code

    13.2GitHub & Project Demo Link

1.INTRODUCTION:

This project “Car Resale Value Prediction” aims to build a model to predict used cars' reasonable prices based on multiple aspects, including vehicle mileage, year of manufacturing, fuel consumption, transmission, road tax, fuel type, and engine size. This model can benefit sellers, buyers, and car manufacturers in the used cars market. Upon completion, it can output a relatively accurate price prediction based on the information that users input. The model building process involves machine learning and data science. The dataset used was scraped from listings of used cars. Various regression methods, including linear regression, polynomial regression, support vector regression, decision tree regression, and random forest regression, were applied in the research to achieve the highest accuracy. Before the actual start of model-building, this project visualised the data to understand the dataset better. The dataset was divided and modified to fit the regression, thus ensuring the performance of the regression.

1.1Project 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 data set. This ability to capture data, analyze it and use it to personalize a shopping experience or implement is the future of retail.

Parameters involved :

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

1.2Purpose:

Car resale value prediction helps the user to predict the resale value of the car depending upon various features like kilo-meters driven, fuel type, etc. This resale value prediction system is made for general 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.

2.LITERATURE SURVEY

2.1Existing problem

The problem is defined as the optimized way to estimate insurance cost based on the manufacturer with some additional costs incurred by the Government in the form of taxes. As the existing methods for estimating the cost takes a lot of time and energy and due to the increased price of new cars and the inability of customers to buy new cars due to the lack of funds, used cars sales are on a global increase.

The prices of new cars in the industry is fixed by the So, customers buying a new car can be assured of the money they invest to be worthy. There is a need for a used car price prediction system to effectively determine the worthiness of the car using a variety of features. Even though there are websites that offer this service, their prediction method may not be the best. Besides, different models and systems may contribute to

predicting power for a used car’s actual market value. It is important to know their actual market value while both buying and selling.

2.2 References

[1] Kanwal Noor, 2017, Vehicle Price Prediction System using Machine Learning Techniques International Journal of Computer Applications. Volume 167 - Number

[2] Mariana Lusitania et al, (2009). Support vector regression analysis for price prediction in a vehicle leasing application [3] Richardson, M. S. (2009). Determinants of used vehicle resale value.

[3] Listiani, M. (2009). Support vector regression analysis for price prediction in a car leasing application (Doctoral dissertation, Master thesis, TU Hamburg-Harburg).

[4]T. D. Phan, "Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City Australia", 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), pp. 35-42, 2018.

[5]K. Samruddhi and R. Ashok Kumar, "Used Car Price Prediction using K-Nearest Neighbor Based Model", International Journal of Innovative Research in Applied Sciences and Engineering, vol. 4, no. 3, pp. 686-689, 2020. [6]O. Celik and U. O. Osmanoglu, "Prediction of The Prices of SecondHand Cars", Avrupa Bilim ve Teknoloji Dergisi, no. 16, pp. 77-83, Aug. 2019

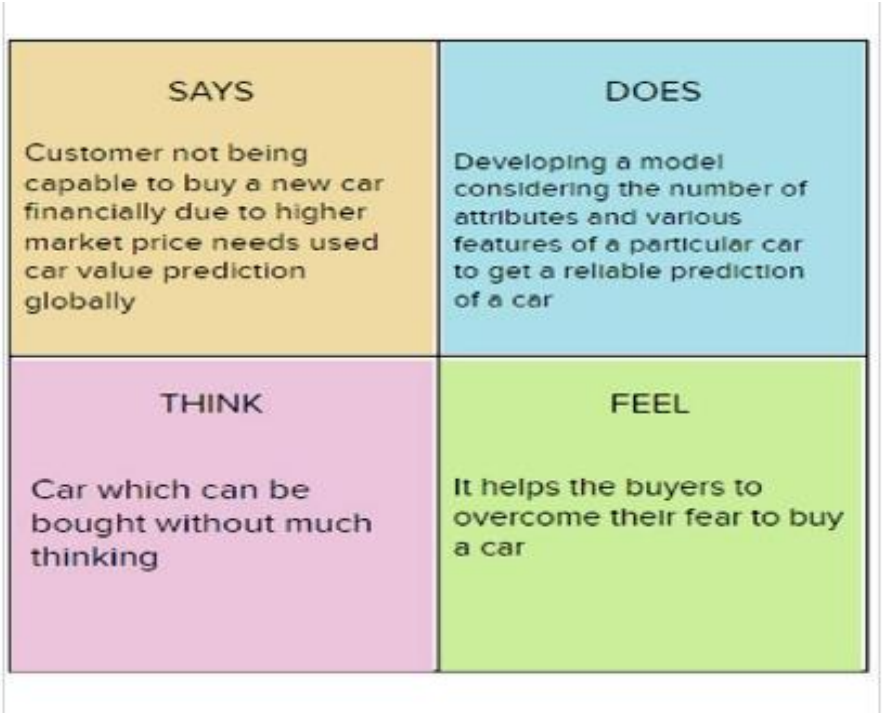
2.3 Problem Statement Definition

Due to the huge requirement of used cars and lack of experts who can determine the correct valuation . A Effective solution to predict used cars prices by scraping data from websites that sell used cars, and analyzing the different aspects and factors that lead to the actual used car price valuation.

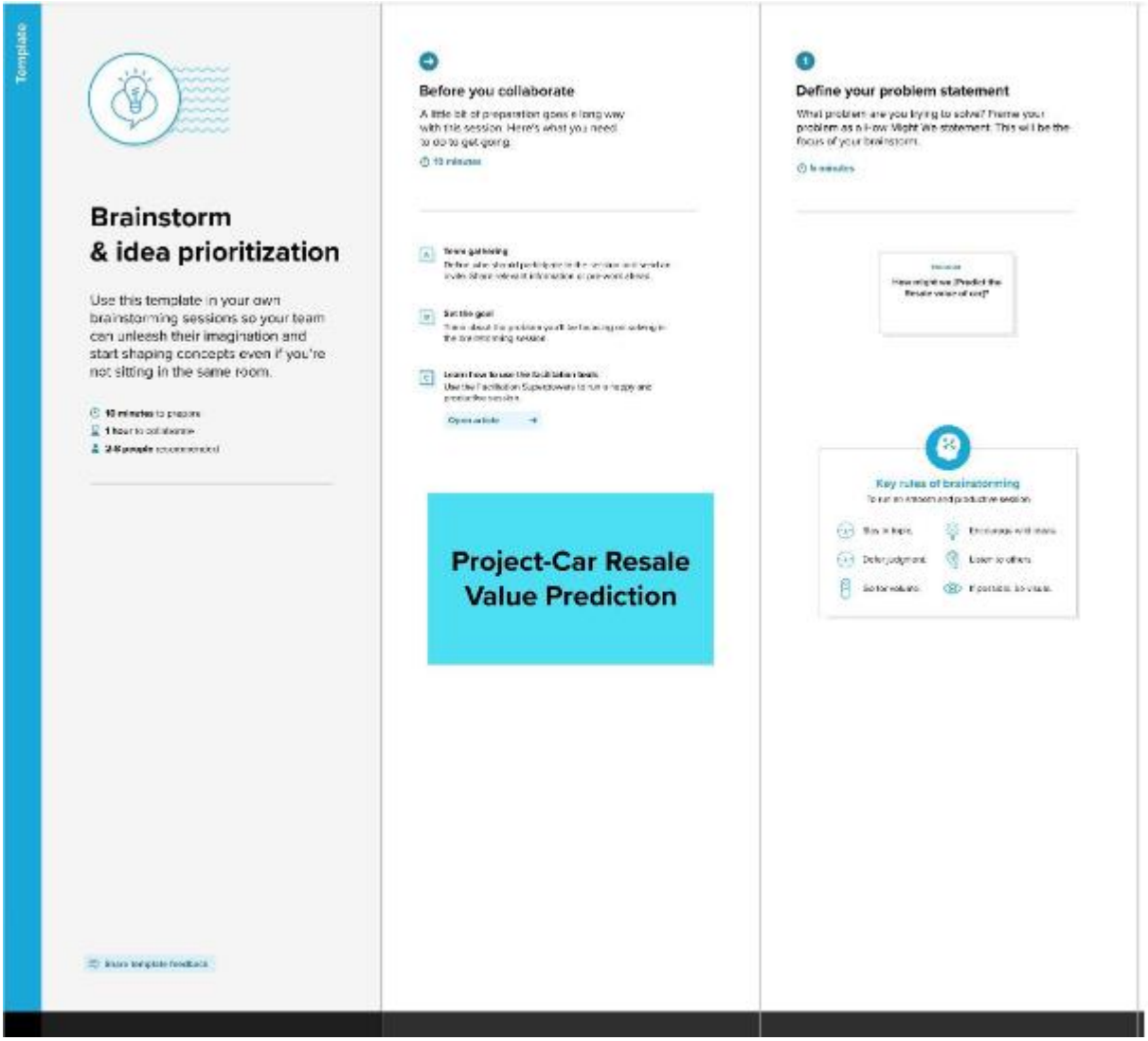
- ✓ to help guide the individuals looking to buy or sell cars and to give them a better insight into the automotive sector
- ✓ Therefore, to help consumers avoid falling victims to some dealer, this car resale value prediction hopes to equip consumers with right tools to guide them in their shopping experience
- ✓ Another goal of the project is to explore new methods to evaluate used cars prices and to compare their accuracy

3.IDEATION & PROPOSED SOLUTION

3.1Empathy Map Canvas



3.1Ideation & Brainstorming



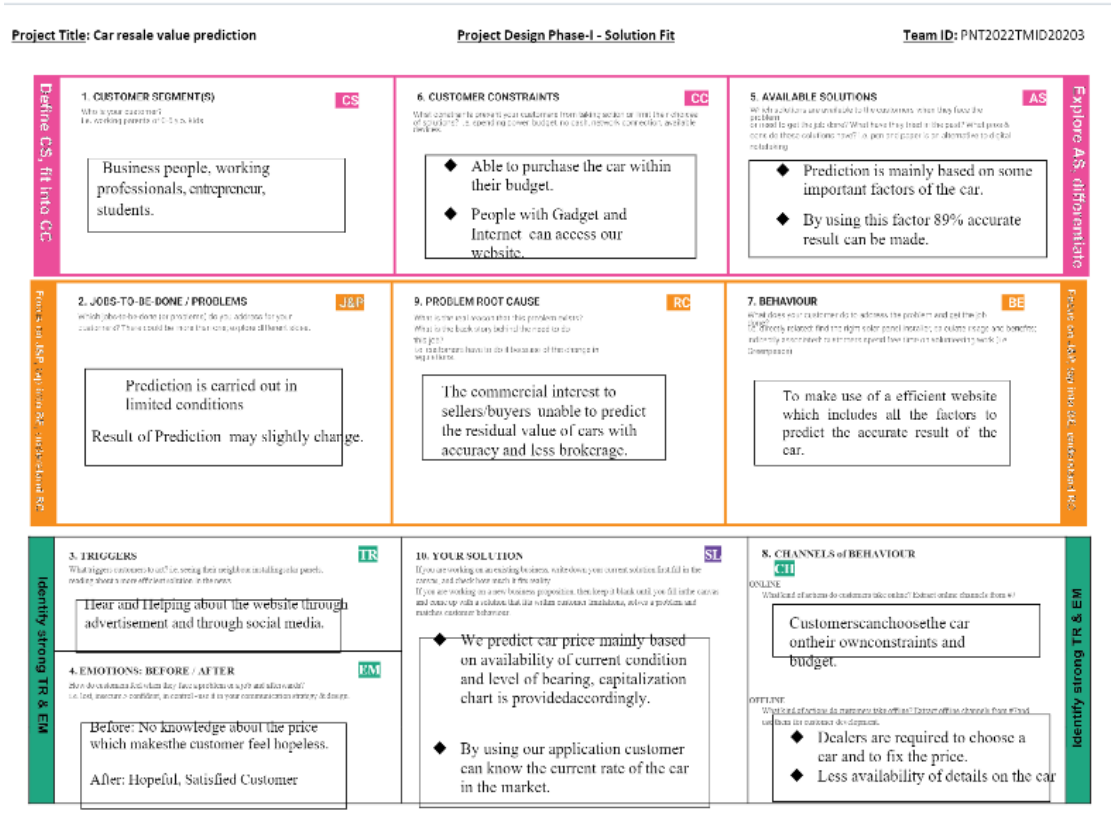
3.3 Proposed Solution

. No.	arameter	)escription

.	roblem Statement (Problem to be solved)	o Predict the value of the sed car using Data Science
.	dea / Solution lescription	re-owned car sale is more opular in developing ountry. People Who plan to urchase used car often truggle to find a within a udget as well as to predict ne price of the second-hand ar. So our project helps a otential buyer to estimate t rice of a Second-hand car. nalysis Data using various Machine learning Algorithm
.	Noovelty / Uniqueness	Ve predict Used Car price ainly based on the car ondition and mitigation lev f quality, capitalization cha s provided accordingly.
.	Social Impact / Customer Satisfaction	By Using this application Customer can know the resale price of car in the market and chart provided user to know good maintenance and make quality of car.
.	Business Model (Revenue Model)	Dealing with mitigation measure makes our idea futuristic and we provided detail information through chart. Being clear and unique, it attracts more customer leading higher revenue.

.	Scalability of the olution	Whatever may be the vehic type or count of vehicle, th system predicts the appropriate resale value If Multiple user access the system at same time, it process scalable.

3.4 Problem Solution fit



4 . REQUIREMENT ANALYSIS

4.1 Functional requirement

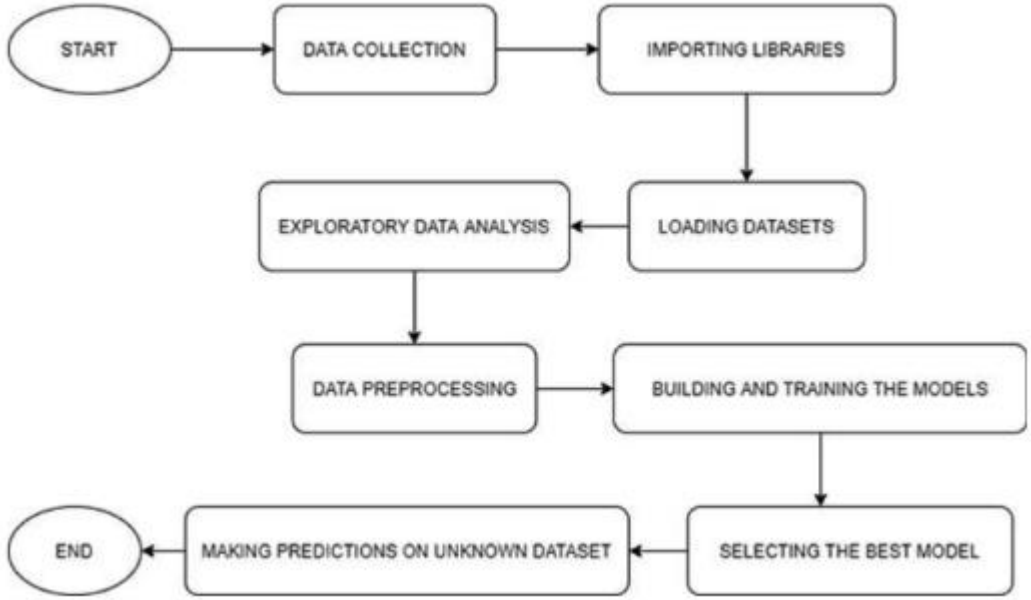
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	• Registration through Form • Registration through Gmail • Registration through Instagram • Registration through Facebook
FR-2	User Confirmation	• Confirmation via Email • Confirmation via Message
FR-3	User Profile	• View account details • Change password
FR-4	Car details	• Adding the new car • Getting car details • View and update details
FR-5	Maintain database	• Store car details in car database • Store user details in user database
FR-6	Value prediction	• Predict the value of the resale car using the model and details entered • Display the predicted value
FR-7	Feedback	• Feedback through form

Non-functional Requirements:

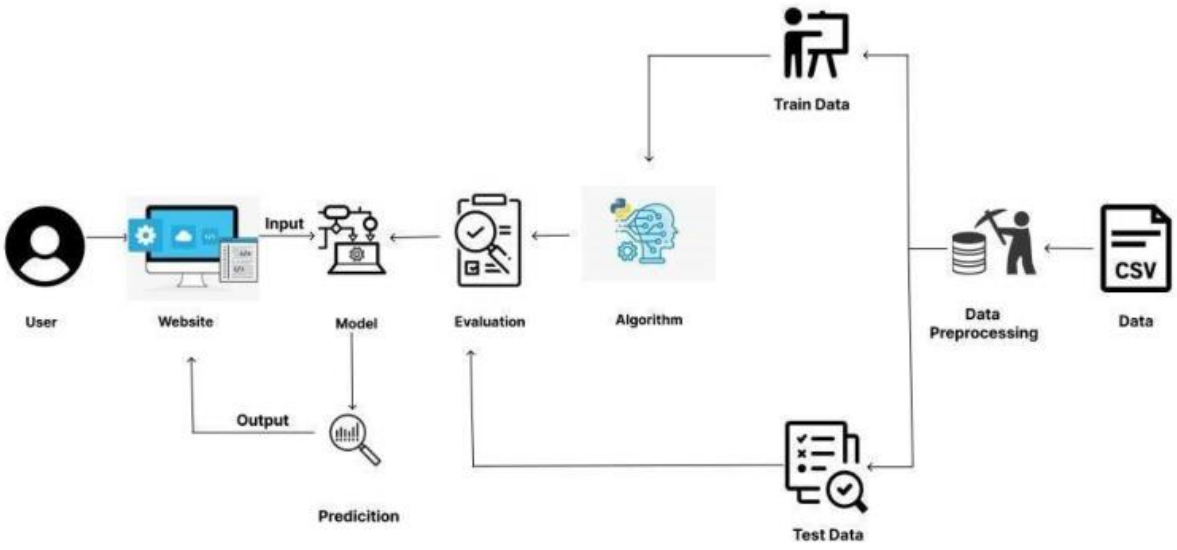
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	• User friendly UI • Clear instructions and sample examples will be provided • Easy process flow to predict value
NFR-2	**Security**	• User authentication while entering website • No information is shared with third party • User can see only his details
NFR-3	**Reliability**	• Data will be stored and replicated so that data loss can be avoided • Rate of occurrence of failure is very less
NFR-4	**Performance**	• Quick prediction results • Fast website loading • Efficient ML algorithm to provide accurate result with less time complexity
NFR-5	**Availability**	• Application can be accessed from both mobile and desktop • Single page failure does not affect the whole website • Uninterrupted user services
NFR-6	**Scalability**	• Able to handle large amount of data and traffic globally without failure • Database can be scaled according to the usage in a cost effective manner

5.PROJECT DESIGN

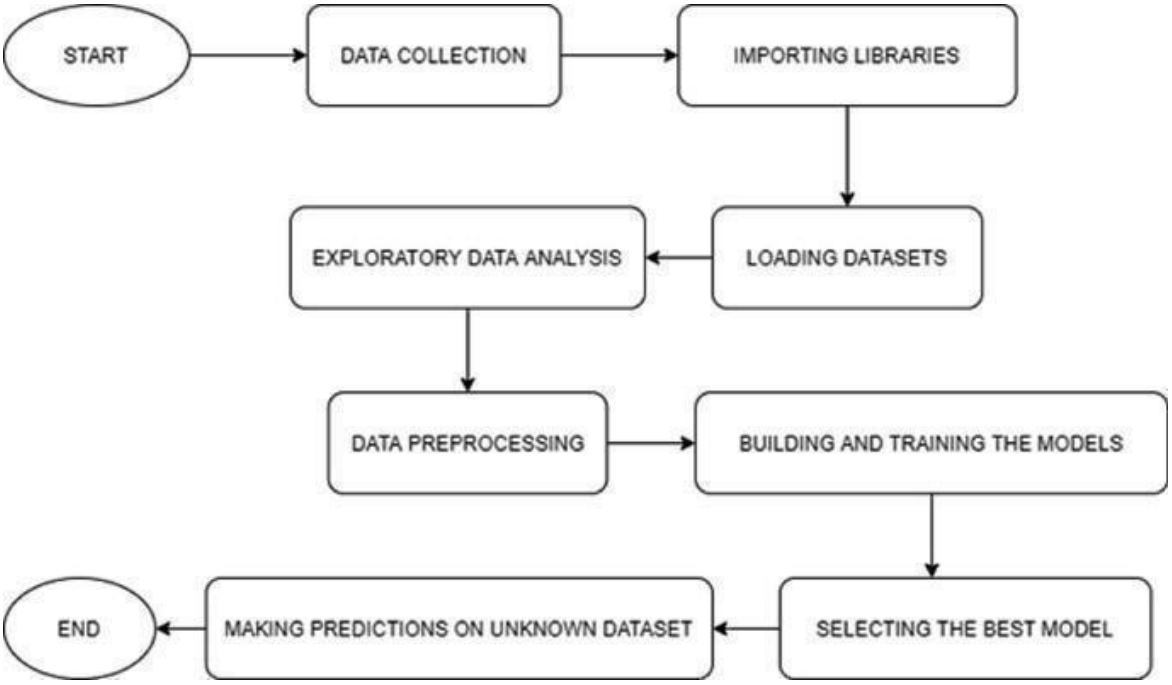
5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture



5.3 User Stories:



User Stories

Use the below template to list all the user stories for the product.

ser Type	unctional equirem t (Epic)	ser Story umber	ser Story / Task	cceptance criteria	riority	elease
ustomer (ser)	want to uy a used ar	SN-1	s a user, I can register for th pplication by entering my mail, password, and onfirming my password.	can access my account ashboard	igh	print-1
		SN-2	s a user, I will rece onfirmation email once I ha egistered for the application	can receive confirmati mail & click confirm	igh	print-1
		SN-3	s a user, I can register for th esource want	can access the resource nd now about the car arieties and their mode nd value of he car	ow	print-2
		SN-4	s a user, I can register for th pplication through Gmail		Iedium	print-1
	ogin	SN-5	s a user, I can log into the pplication by entering email assword		igh	print-1
	ashboard					
ustomer Web user)						
ustomer Ca xecutive						
dministrato						

6 .PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Sprint	l Requirement (Epic)	er Story umber	ser Story / Task	tory Poi	riority	eam Members

print-1	)ata Entry	ISN-1	.s a user, I can enter the ca :esale details in the pplication.		Medium	.alaiselvi P
print-1	)btain output	ISN-2	.s a user, I will receive car :esale cost in the applicatio		ligh	Mary Shipani J
print-1	)ata Entry	ISN-3	.s a user, I can enter the ca :esale details in the pplication.		Medium	henbagalakshmi S

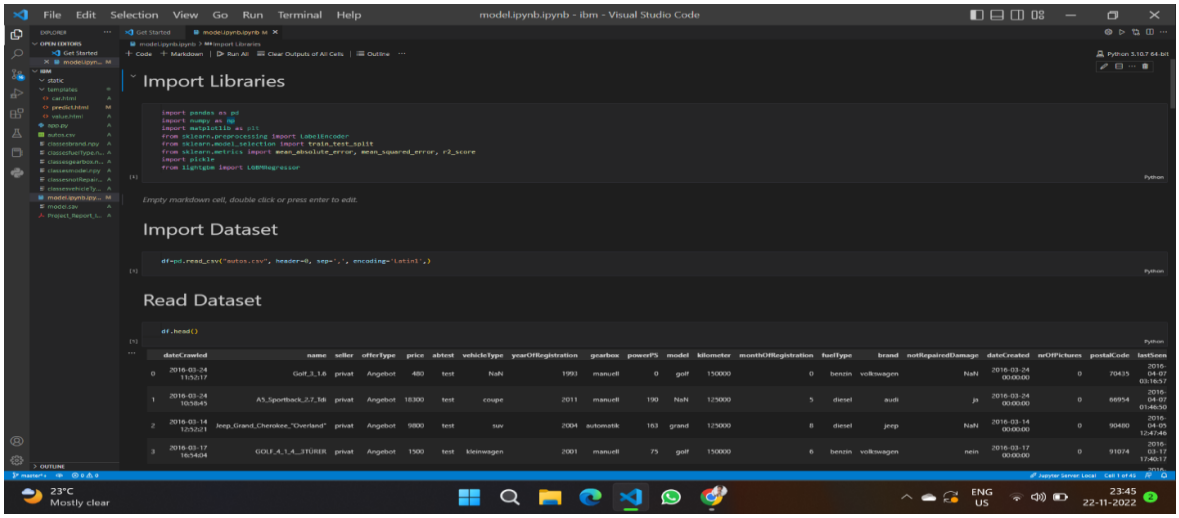
print-1	Obtain output	USN-4	As a user, I will receive ar resale cost in the pplication	2	High	riyadharshini G
print-2	ogin	USN-5	As a user, I can log into he application by ntering email & assword	1	ow	henbagalakshmi
print-2	Dashboard	USN-6	As a user, I can log into he dashboard by ntering sername & password	2	High	Mary Shipani J
print-3	Car resale value rediction	USN-7	As a user, I can access he car resale value rediction section	3	High	Kalaiselvi P
print-3	Customer queries	USN-8	As a customer care xecutive, I can check he customer queries th osted in the ebsite.	2	High	henbagalakshmi
print-4	Maintaining website	USN-9	As an administrator, I an maintain website an nhance he online presence.	1	Medium	riyadharshini G
print-4		USN-10	As an administrator, I an maintain issues in nalysing alues.	1	Medium	Mary Shipani J
print-3		USN-11	As an administrator, I an update the website ontent	2	High	Kalaiselvi P
print-4		USN-12	As an administrator, I an improve the websit	2	High	riyadharshini G

6.2Sprint Delivery Schedule

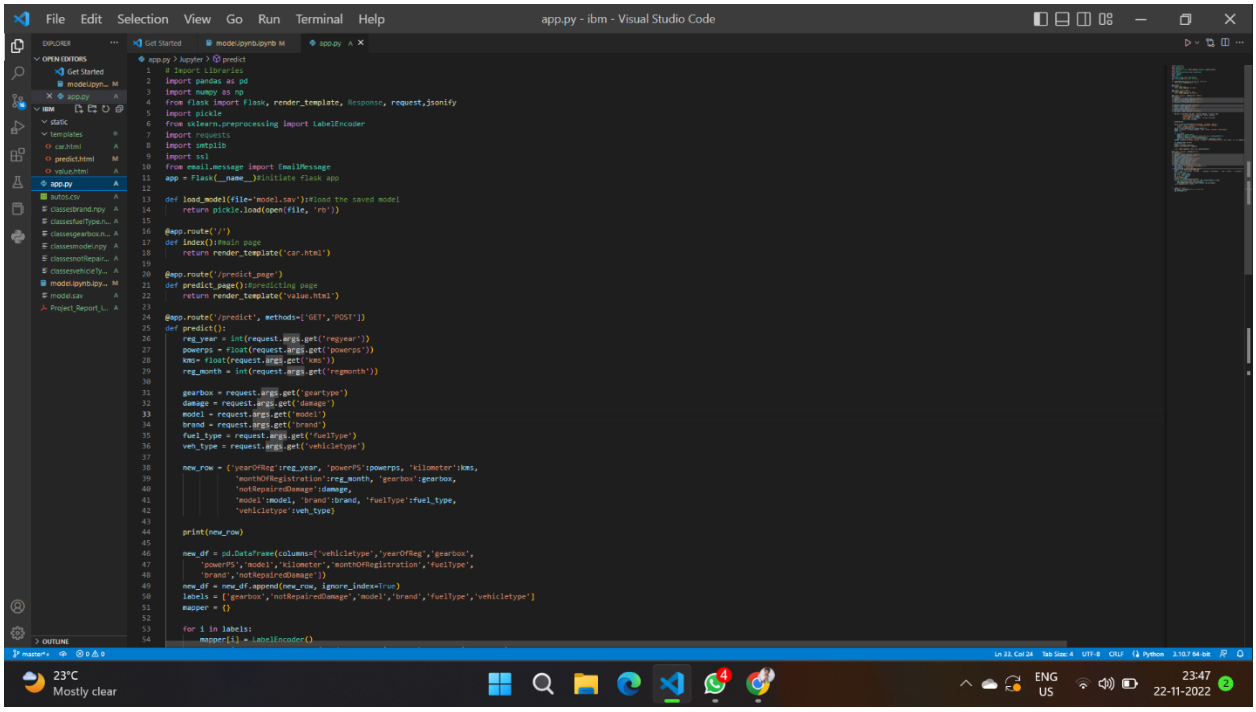
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022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

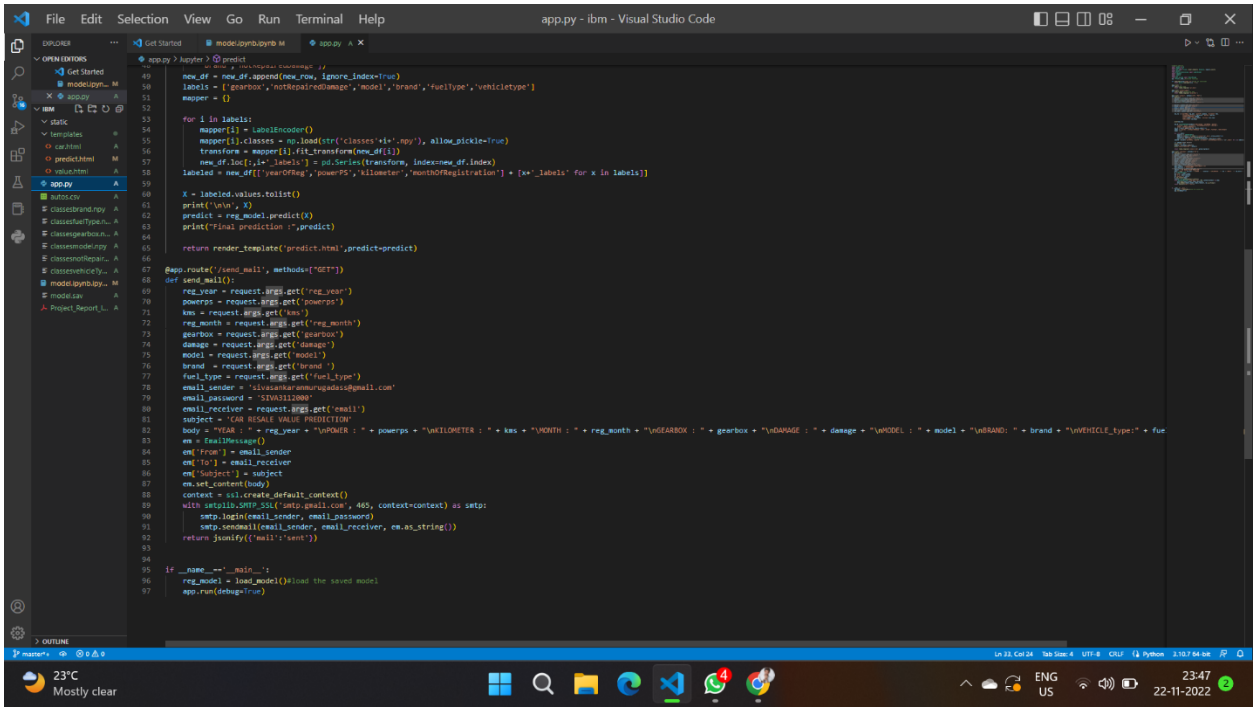
7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature 1



7.2 Feature 2





8.TESTING

8.1 Test Cases

DataEntryPag TC_002			Func nal			Data ntry Pag	er is able to enter alues		1.Enter URL o 2.Verify all ements display .Press the Chec utton in the hor erifyall the UI isplayed or not erify if all valu ntered
DataEntryPag TC_003			Func nal			a Entry age	heOutput Display age can be reachab		1.Enter URL o 2.Verify all ements display .Press the Chec utton in the hor erifyall the UI isplayed or not erify if all valu ntered ress thesubmit
OutputDispla age_TC_001			I			Output DisplayP e	Verify all the U elements in Output Display pag rendered properly		1.Enter URI o 2.Verify all lements display .Press the Chec utton in the hor

									erifyall the UI isplayed or not erify if all valu ntered ress thesubmit erifyall the UI isplayed or not
OutputDispla age_TC_002			Func nal			Output isplayP e	r is ableto get redicted result		1.Enter URI o 2.Verify all lements display .Press the Chec utton in the hor erifyall the UI isplayed or not erify if all valu ntered ress thesubmit erifyall the UI isplayed or not erifyif the prec alueis displaye

Test Scenarios :

Verify user is able to see home page?

Verify user is able to navigate to data entry page?

Verify user is able to see data entry page?

Verify user is able to enter values in the fields?

Verify user is able to navigate to output display page?

Verify user is able to view the output display page?

8.2 User Acceptance Testing

Defect Analysis

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
y Design	10	4	2	3	20
uplicate	1	0	3	0	4
xternal	2	3	0	1	6
ixed	11	2	4	20	37
lot Reproduced	0	0	1	0	1
kipped	0	0	1	1	2
Von't Fix	0	5	2	1	8
otals	24	14	13	26	77

Test Cast Analysis

ection	Total Cases	Not Tested	Fail	Pass
rint Engine	7	0	0	7
lient Application	51	0	0	51
ecurity	2	0	0	2
utsource Shipping	3	0	0	3
xception Reporting	9	0	0	9
inal Report Output	4	0	0	4
ersion Control	2	0	0	2

9. RESULTS

9.1 Performance Metrics

{'mae': 1235.112086905962,
'mse': 9377053.62710202,
'rmse': 3084.6815065692977,
'rmsle': 8.43744027403009,
'r2': 0.8361221626879432,
'adj_r2_score': 0.8261152969113608}

The model is tested with the various damaged car images which is not used during the training and validation of the model which also shows that the model works with the accuracy of about 98% in the overall performance

10. ADVANTAGES & DISADVANTAGES

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

CONS:

- Less effective

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.

12.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

13.APPENDIX

13.1 SOURCE CODE

```
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='xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx',
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)                                    17

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',inplac
e=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:].value
```

```
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'))
```

13.2 GitHub Repo:

https://github.com/IBM-EPBL/IBM-Project-24882-1659950473

13. 2 Demo Link:

https://drive.google.com/file/d/1lPCXGQdyOUUwyj9WIBEuNYzfzp9eiucx/view?usp=share_link