A PROJECT REPORT ON CAR RESALE VALUE PREDICTION

DOMAIN : Applied Data Science

TEAM :IBM-Project-35390-1660284416

COLLEGE NAME : Government College of Engineering-Bargur

TEAM MEMBERS

TEAM ID: PNT2022TMID06615

TEAM LEADER : GAJALAKSHMI R(61071911114)
TEAM MEMBER 1: SAVITHA R(61071911144)
TEAM MEMBER 2: RAMYA S (61071911136)
TEAM MEMBER 3: SOWFIYA BEGAM S(61071911146)

TABLE OF CONTENT

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

In this project we have used different algorithms with different techniques for developing Car resale value prediction systems considering different features of the car. In a nutshell, car resale value prediction helps the user to predict the resale value of the car depending upon various features like kilometre driven, fuel type, etc. Need for the System This resale value prediction system is made for general purpose to just predict the amount that can be roughly acquired by the user. We try to predict the amount of resale by best 70% accuracy so the user can get estimated value before he resales the car and doesn't make a deal in loss. 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.

BACKGROUND

In this project, we mainly focus on the analysis of the Vehicle Resale Predict and then predict the results through them using training data. AI calculations can be utilized to anticipate the retail worth of a vehicle, in light of a specific arrangement of highlights. Various sites have various calculations to create the retail cost of the trade-in vehicles, and subsequently there is certainly not a brought together calculation for deciding the cost. Via preparing measurable models at foreseeing the costs, one can undoubtedly get a good guess of the cost without really entering the subtleties into the ideal site.

1.2PROJECT PURPOSE

The main idea of making a car resale value prediction system is to get hands-on practice for python using **Data Science**. Car resale value prediction is the system to predict the amount of resale value based on the parameters provided by the user. User enters the details of the car into the form given and accordingly the car resale value is predicted.The system is defined in the python language that predicts the amount of resale value based on the given information. The system works on the trained dataset of the machine learning program that evaluates the precise value of the car. User can enter details only of fields like purchase price of car, kilometres driven, fuel of car, year of purchase.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. The new system developed by us consists of two parts - Data gathering and Prediction using **Machine Learning based algorithms**.

CHAPTER 2

LITERATURE SURVEY

The first paper is Predicting the price of Used Car Using Machine Learning Techniques. In this paper, they investigate the application of supervised machine learning techniques to predict the price of used cars in Mauritius. The predictions are based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. The Second paper is Car Price Prediction Using Machine Learning Techniques. Considerable number of distinct attributes are examined for the reliable and accurate prediction

2.1 EXISTING PROBLEM

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	Car dealer	sell used cars	Increasing price	rising interest rates, tariffs, and energy concerns, car dealerships are expected to have fewer sales, especially with newer vehicles	Sad and Worried
PS-2	Common People	Buy 2nd hand cars	Can't decide on cars	Too many car models & prices. Common man can't decide correct car.	Fear of making wrong choice
PS-3	Budget Oriented People	Buy cheap cars	Price is not justified	Price is increased and cannot justify price	Paying more for cars
PS-4	Seller	To sell my car at reasonable price	Deciding on the price is hard	Too many complications in calculating the correct price for selling cars	Unhappy for not selling car at correct price.

2.2 REFERENCES

S.NO	AUTHORS	PAPER	YEAR
1	Sameerchand Pudaruth	Predicting the Price of Used Cars using Machine Learning Techniques	IJICT 2014
2	Enis gegic, Becir Isakovic Dino Keco Zerina Masetic Jasmin Kevric,	Car Price Prediction Using Machine Learning	TEM Journal 2019
3	Ning sun Hongxi Baily uxia Geng Huizhu Shi	Price Evaluation Model In Second Hand Car System Based On BP Neural Network Theory	Hohai University Changzhou, China
4	Nitis Monburinon Prajak Chertchom Thongchai Kaewkiriya Suwat Rungpheung Sabir Buya Pitchayakit Boonpou,	Prediction of Prices for Used Car by using Regression Models	ICBIR 15
5	Doan Van Thai Luong Ngoc Son Pham Vu Tien Nguyen Nhat Anh Nguyen Thi Ngoc Anh	Prediction car prices car prices using qualify qualitative data and knowledge-based system	Hanoi National University

2.3 PROBLEM STATEMENT

DEFINITION

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. Location like Delhi, Chennai, Mumbai. It based on year of manufacturing like 2020, 2022, type of fuel namely Petrol, Diesel. Price range or Budget and type of transmission which the customer prefers like Automatic or Manual, Mileage to name a few characteristic features required by the customer. The project Car Price Prediction deals with providing the solution to these problems. Car makers face several challenges in the second-hand market. The depth crisis in the European Union, the general problem of overcapacity, increasing competition from Asian manufacturers, and the trend toward more eco-friendly cars are only a few factors that add to the difficulty of selling used vehicles in the second-hand market and decrease sales margins. Therefore, car makers require sophisticated decision support systems to sustain the profitability of the used car business. A core component of such

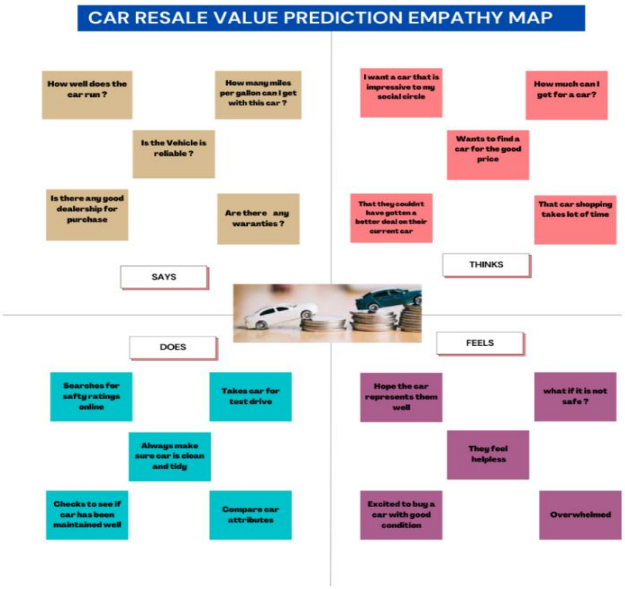
systems is a prediction model that estimates resale prices on the basis of car attributes and other factors.

CHAPTER 3

IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP

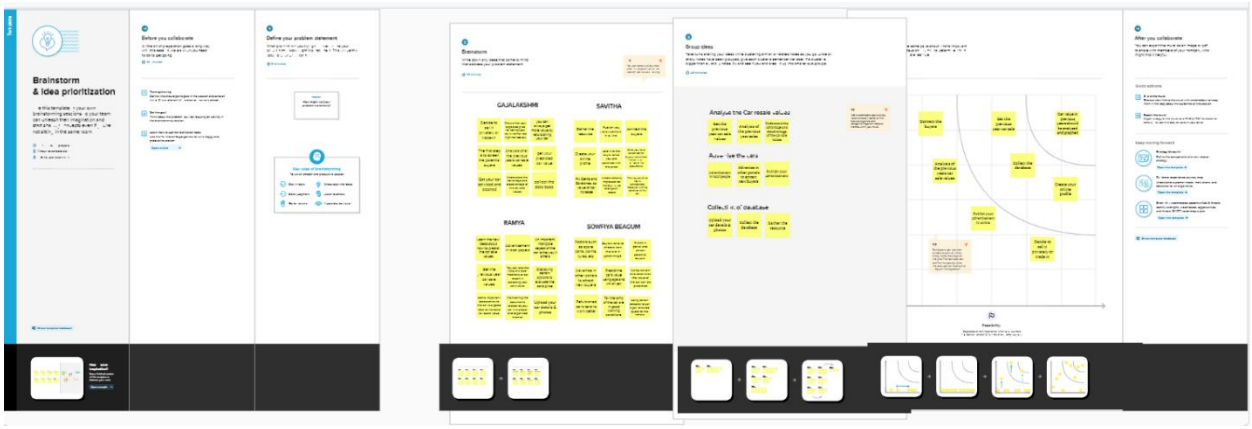
An empathy map is a visualization tool used to articulate what a product team knows about a user. This tool helps product teams build a broader understanding of the "why" aspect behind user needs and wants. This tool forces product teams to practice empathic design, which shifts the focus from the product they want to build to the people who will use this product. The four empathy map quadrants look at what the user **says**, **thinks**, **feels**, and **does**.



3.2 IDEATION AND BRAINSTROM

The term brainstorming has now become the accepted way of referring to group attempts to solve specific problems or develop new ideas by amassing spontaneous, unrestrained contributions by members. Brainstorming can be used to generate possible solutions for simple problems, but it is unrealistic to expect it to accomplish most problem-solving or planning tasks. The technique is of value as part of a larger effort that includes individual generation of information and ideas and subsequent compilation, evaluation, and selection. Brainstorming

can be used to generate components of a plan, process, solution, or approach and to produce checklists.

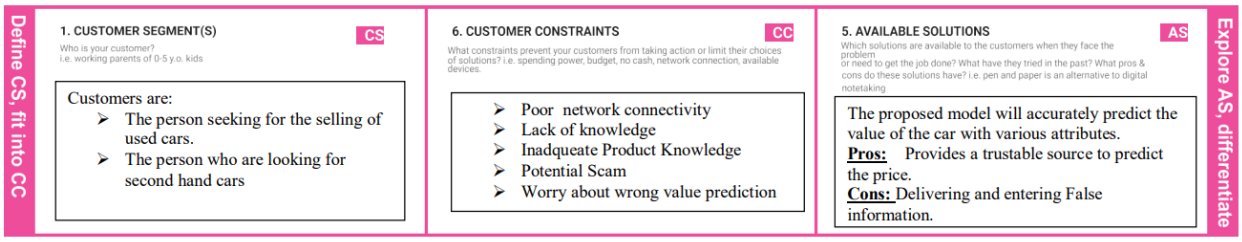


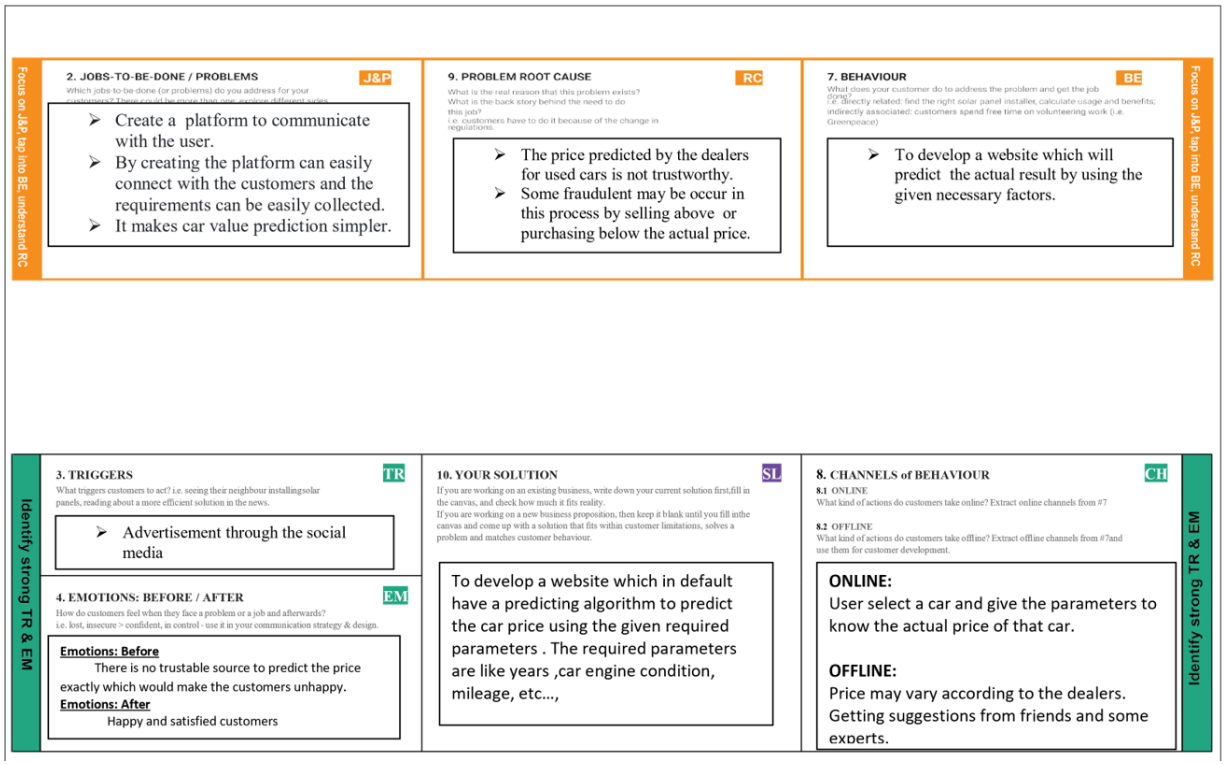
3.3 PROPOSED SOLUTION

S.NO	Parameter	Description
1	Problem statement	The main aim of project is to predict the resale value of a used car using linear regression algorithms.
2	Idea/ solution description	The resale value of car depends on km, fuel etc The data is then pre-processed to handle missing values and outliers and split into depend variable and independent variable. After that the model is developed using regression algorithms to predict the resale price to the car.
3	Novelty/Uniqueness	This is a real time problem which can benefit both customer and seller. The novelty of this proposal is to predict the resale value as near as possible to actual value.
4	Social impact	Provided the current economic times it is more likely that the user age of second-hand cars will increase. This mutual commercial interest to both customer and sellers can increase profit.
5	Business model	The proposed solution could be sold to reseller so that they could use to find perfect price for bidding. It could be developed into an application and get revenue from it if more on of users started to using it

		to find the best value to second hand cars.

3.4 PROBLEM SOLUTION FIT





CHAPTER 4

REQIUREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

Followingarethefunctionalrequirementsoftheproposedsolution.

FRNo.	FunctionalRequirement(Epic)	SubRequirement(Story/Sub-Task)
FR-1	UserRegistration	RegistrationthroughWebsite
FR-2	UserConfirmation	Confirmationviawebsite
FR-3	CarRegistration	Registeringthecardetails
FR-4	ValuePrediction	Predictingthecarresalevalue

4.2 NON-FUNCTIONAL ANALYSIS

Followingarethenon-functionalrequirementsoftheproposedsolution.

FRNo.	Non-FunctionalRequirement	Description

NFR-1	Usability	Predictingtheresalevalue
NFR-2	Security	Providingsecuritytothewebsite
NFR-3	Reliability	Providinghighreliabilitybypredictingvaluesfordifferenttypesofcars
NFR-4	Performance	Providinghighperformance by usingsomemachinelearningtechniques
NFR-5	Availability	Itisusedforalltypesofcars
NFR-6	Scalability	Predictingvaluesfordifferenttypesofcars

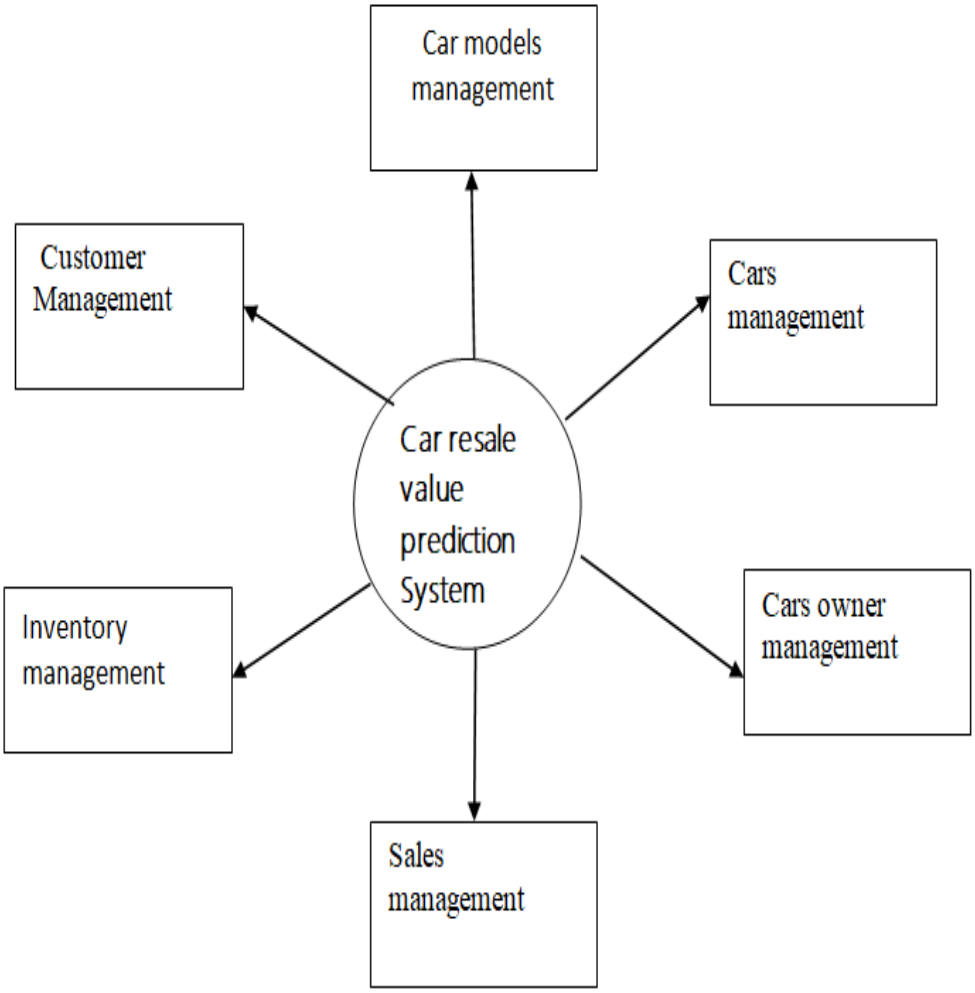
CHAPTER 5

PROJECT DESIGN

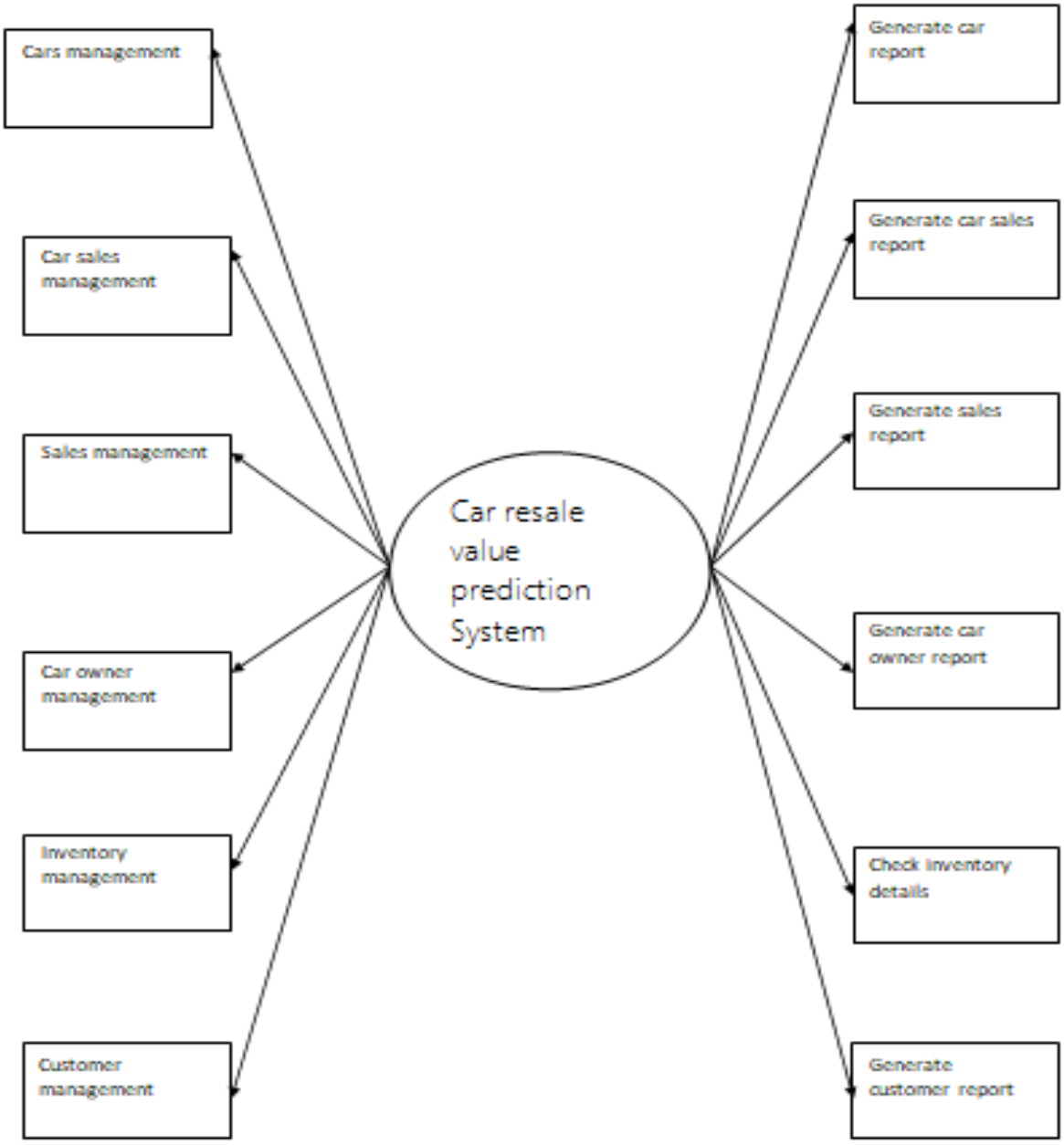
5.1 DATA FLOW DIAGRAM

ADataFlowDiagram(DFD)isatraditionalvisualrepresentationofthe informationflowswithina system. A neat and clear DFD can depict the right amount of the system requirementgraphically. It shows how data enters and leaves the system, what changes the information,andwheredatais stored.

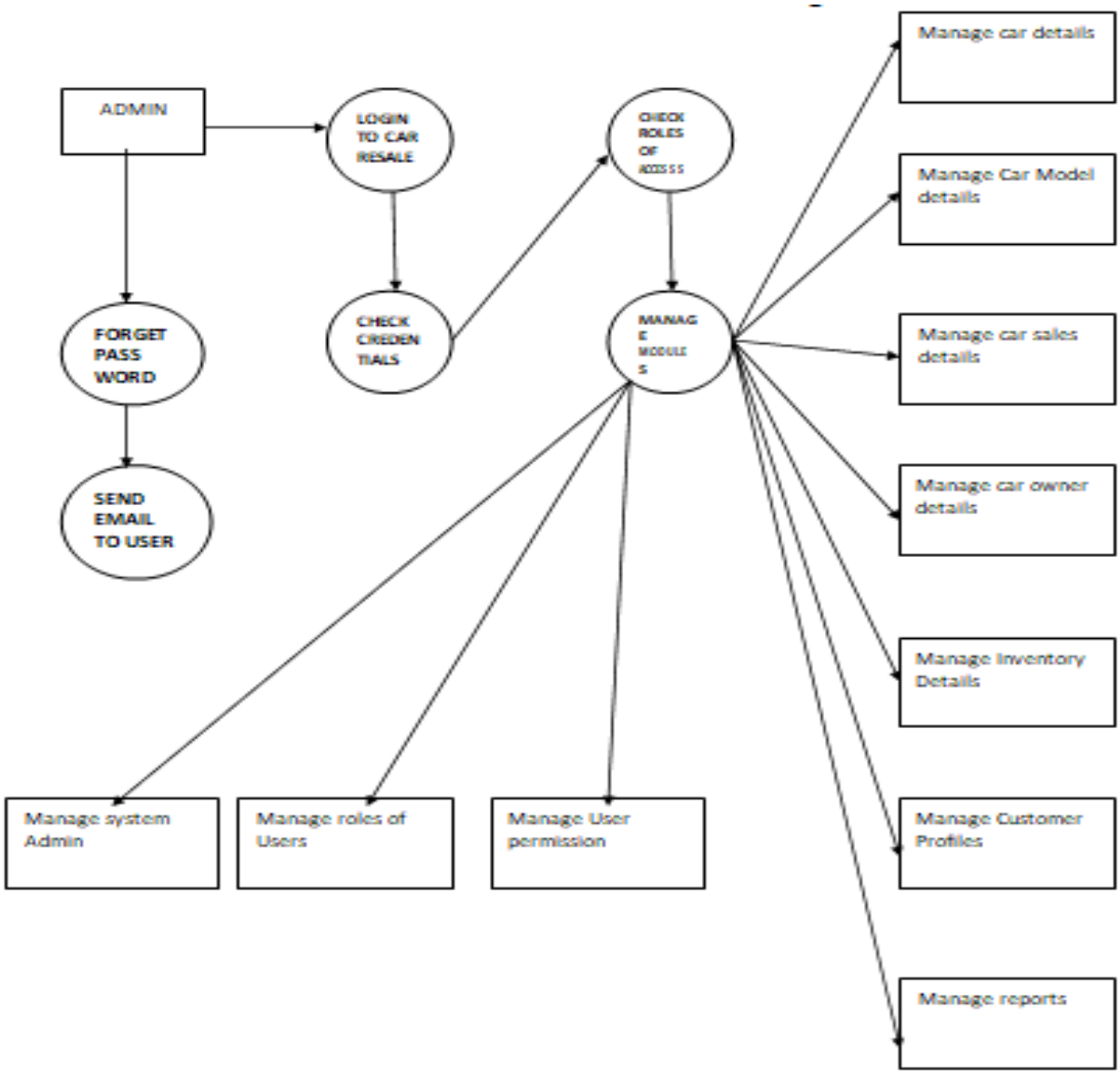
ZeroLevelDataFlowDiagram



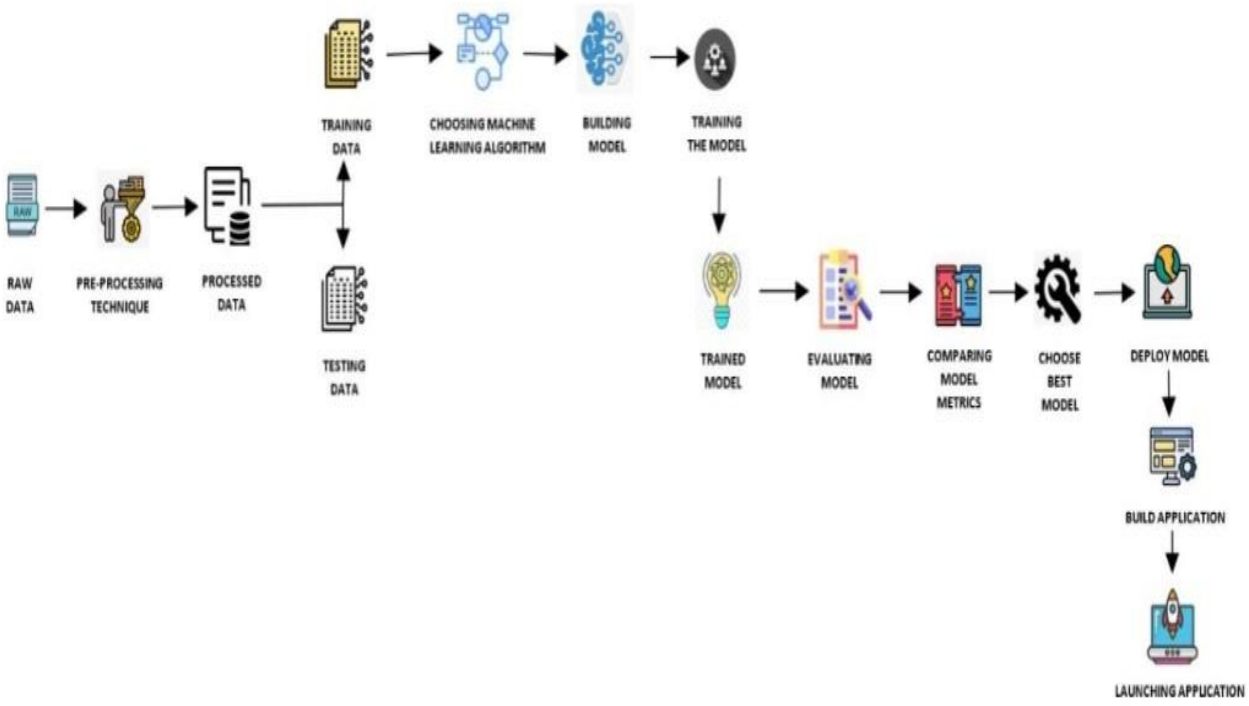
First Level Data Flow Diagram



Second LevelDataFlowDiagram



5.2SOLUTION & TECHNICAL ARCHITECTURE


RAW DATA
PRE-PROCESSING TECHNIQUE
PROCESSED DATA
TRAINING DATA
TESTING DATA
CHOOSING MACHINE LEARNING ALGORITHM
BUILDING MODEL
TRAINING THE MODEL
TRAINED MODEL
EVALUATING MODEL
COMPARING MODEL METRICS
CHOOSE BEST MODEL
DEPLOY MODEL
BUILD APPLICATION
LAUNCHING APPLICATION

CHAPTER 6

PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

USER	SPRINT	FUNCTIONAL REQUIREMENT	USER STORY NUMBER	USER STORY/TASK	PRIORITY	TEAM MEMBER
Admin	Sprint 1	Dataset collection	USN-1	Collect the required data for the car resale prediction	High	Sowfiya Begam S
	Sprint 1		USN-2	Perform data cleaning to optimize the dataset	Medium	Sowfiya Begam S
	Sprint 2	Model building	USN-3	Build the model using regression algorithms classify the data	High	Ramya S
	Sprint 2		USN-4	Deployment of ML model using IBM cloud	High	Ramya S
	Sprint 3	Train the model	USN-5	Integrate the web app developed using flask with IBM model	High	Savitha R
Customer	Sprint 3		USN-6	Details about the application and the car resale process	Low	Savitha R

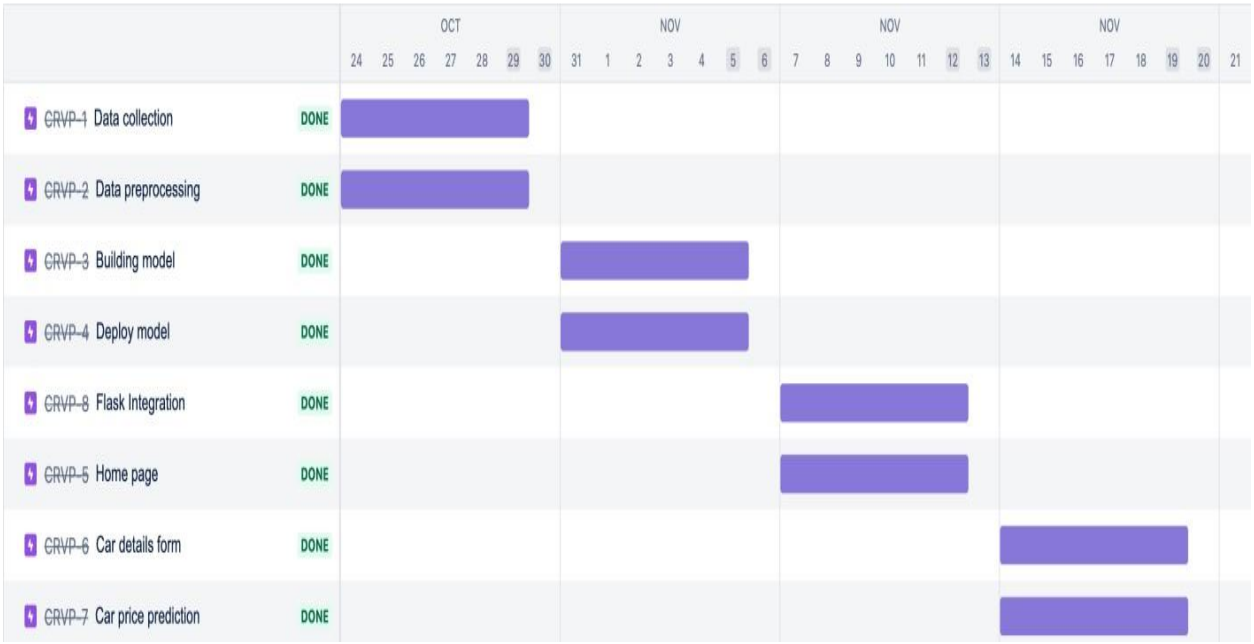
User	Sprint	FUNCTIONAL REQUIREMENT	USER STORY NUMBER	USER STORY/TASK	PRIORITY	TEAM MEMBERS
	Sprint 4	Car details	USN-7	As a user I	Medium	Gajalakshmi R

				should give the car details like car model engine and fuel type.		
	Sprint 4	Car details	USN-8	As a user I can view the current rate of the used car price	High	Gajalakshmi R

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 (Actual)
Sprint-1	10	6 Days	24 Oct 2022	29 Oct 2022	10	29 Oct 2022
Sprint-2	10	6 Days	31 Oct 2022	05 Nov 2022	10	05 Nov 2022
Sprint-3	10	6 Days	07 Nov 2022	12 Nov 2022	10	12 Nov 2022
Sprint-4	10	6 Days	14 Nov 2022	19 Nov 2022	10	19 Nov 2022

6.3 REPORTS FROM JIRA



CHAPTER 7

CODING & SOLUTIONING

7.1 LINEAR REGRESSION

In [23]:
```
lr=LinearRegression()
```

In [24]:
```
pipe=make_pipeline(column_trans,lr)
```

In [25]:
```
pipe.fit(X_Train,Y_Train)
```

Out[25]:
```
Pipeline(steps=[('columntransformer',
                 ColumnTransformer(remainder='passthrough',
                                   transformers=[('onehotencoder',
                                                  OneHotEncoder(categories=[array(['Ambassador CLASSIC 1500', 'Ambassador Classic 2000',
       'Ambassador Grand 1800', 'Audi A4 1.8', 'Audi A4 2.0',
       'Audi A4 3.0', 'Audi A4 30', 'Audi A4 35', 'Audi A4 New',
       'Audi A5 Sportback', 'Audi A6 2.0', 'Audi A6 2.7', 'Audi A6 2.8',
       'Audi A8 4.2', '...
                                                                                array(['CNG', 'Diesel', 'Electric', 'LPG', 'Petrol'], dtype=object),
                                                                                array(['Dealer', 'Individual', 'Trustmark Dealer'], dtype=object),
                                                                                array(['Automatic', 'Manual'], dtype=object),
                                                                                array(['First Owner', 'Fourth & Above Owner', 'Second Owner',
       'Test Drive Car', 'Third Owner'], dtype=object)]),
                                                  ['car_models', 'company_name',
                                                   'fuel', 'seller_type',
                                                   'transmission',
                                                   'owner'])])),
                ('linearregression', LinearRegression())])
```

In [26]:
```
y_pred=pipe.predict(X_Test)
```

In [27]:
```
r2_score(Y_Test,y_pred)
```

Out[27]:
```
0.8949529491096558
```

In [28]:
```
import pickle
```

7.2DEPLOYED THE MODEL IN IBM CLOUD

```
df_data_3 = pd.read_csv(body)
df_data_3.head()
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='wZDmDWGyiQrOEXwAZIKH-rmqaYTbpN8Y1ClESAYBEdU4',
    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 = 'carresalevalueprediction-donotdelete-pr-f0fawdnct29flo'
object_key = 'CAR DETAILS.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 )

car= pd.read_csv(body)
car.head()
car['owner'].unique()
import os, types
import pandas as pd
from botocore.client import Config
import ibm_boto3
```

7.3 USING DEPLOYED MODEL TO PREDICT THE RESULT

```
app.route('/predict', methods=['POST'])
ef predict():
    company_name = request.form.get('company_name')
    car_models = request.form.get('car_models')
    year = int(request.form.get('year'))
    km_driven = int(request.form.get('kilo_driven'))
    fuel = request.form.get('fuel_type')
    seller_type = request.form.get('seller_type')
    transmission = request.form.get('transmission')
    owner = request.form.get('owner')
    # t=[[car_models,company_name,year,km_driven,fuel,seller_type,transmission,owner]]
    # NOTE: manually define and pass the array(s) of values to be scored in the next line
    payload_scoring = {"input_data": [{"fields": ['car_models', 'company_name', 'year', 'km_driven', 'fuel', 'seller_type',
                                                  'transmission', 'owner'], "values":[[car_models, company_name, year, km_driven, fuel, seller_type, transmission, owner]]}]}
    response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/fb824ca5-dfcf-41e1-979f-4a731cd910b5/predictions?version=2022-11-18', json=payload_sc
                                     headers={'Authorization': 'Bearer ' + mltoken})
    prediction = response_scoring['predictions'][0]['values']
    return str(np.round(prediction[00], 2))
```

Used IBM Watson to deploy the model, instead of storing the large modelfile in the local which is not feasible for practical use. This code sends the API request to the deployed model along with the data that the user had entered for which we want to predict the result. After successful prediction the result comes in the form of json which is later parsed and the resale value is obtained.

```python
from flask import Flask, render_template, request
import pandas as pd
import numpy as np
import pickle
import requests

# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "oUHg5RCJT5AVH-CqNzu1fyA067ZbL9NdbCOH3R0DdnIh"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
                                                                                 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]

header = {'Content-Type': 'application/json',
          'Authorization': 'Bearer ' + mltoken}

app = Flask(__name__)
model = pickle.load(open("CarLinearRegressionModel.pkl", 'rb'))
car = pd.read_csv("Cleaned Car.csv")


@app.route('/')
def index():

    car_models = sorted(car['car_models'].unique())
    company_name = sorted(car['company_name'].unique())
    year = sorted(car['year'].unique(), reverse=True)
    km_driven = sorted(car['year'].unique())
    fuel = sorted(car['fuel'].unique())
```

7.5 BOOTSTRAP

templates > <> index.html > ...

```html
<meta content="width=device-width, initial-scale=1.0" name="viewport">

<title>Car Resale Value Prediction</title>


<!-- Favicons -->


<!-- Google Fonts -->
<link href="https://fonts.googleapis.com/css?family=Open+Sans:300,300i,400,400i,600,600i,700,700i|Jost:300,300i,400,400i,500,500i,600,600i,700,700i|Poppins:300,300i,400,400i,500,500i,600,600i,700,700i" rel="stylesheet">

<!-- Vendor CSS Files -->
<link href="static/vendor/aos/aos.css" rel="stylesheet">
<link href="static/vendor/bootstrap/css/bootstrap.min.css" rel="stylesheet">
<link href="static/vendor/bootstrap-icons/bootstrap-icons.css" rel="stylesheet">
<link href="static/vendor/boxicons/css/boxicons.min.css" rel="stylesheet">
<link href="static/vendor/glightbox/css/glightbox.min.css" rel="stylesheet">
<link href="static/vendor/remixicon/remixicon.css" rel="stylesheet">
<link href="static/vendor/swiper/swiper-bundle.min.css" rel="stylesheet">

<!-- Template Main CSS File -->
<link href="static/css/style.css" rel="stylesheet">
```

CHAPTER 8
TESTING

8.1TEST CASES

Test case ID	Feature Type	Component	Test Scenario	Steps To Execute
HomePage_TC_OO 1	Functional	Home Page	Verify user is able to the home page by clicking on the link	1.Enter URL and click go 2.Verify home page displayed or not
HomePage_TC_OO 2	UI	Home Page	Verify the UI elements are responsive	1.Enter URL and click go 2.Repeat the step 1 in different devices 3.Verify all the Elements are visible and accesible
HomePage_TC_OO 3	Functional	Home page	User is able to click on go to predict screen button	1.Enter URL and click go 2.Click on Go to predict Screen button
PredictionPage_TC _OO1	UI	Prediction page	Verify the UI elements are responsive	1.Enter URL and click go 2.Click on Go to predict Screen button 3.Check all the UI elements are accessible and visible 4.Repeat the steps in different devices
PredictionPage_TC _OO2	Functional	Prediction page	Verify user cannot enter null values	1.Enter URL and click go 2.Click on go to predict screen button 3.Submit the form without entering values
PredictionPage_TC _OO3	Functional	Predictoin page	Verify user cannot enter Invalid values (e.g a string for year field)	1.Enter URL and click go 2.Click on go to predict screen button 3.Enter 'Year' in year field, 'ps' in Power ps field and 'KM' in Kilometers field
PredictionPage_TC OO4	Functional	Prediction page	Verify user is able to get the prediction result on screen	1.Enter URL and click go 2.Click on go to predict screen button 3.Enter the valid values in the form fields 4.Click on submit button

8.2USER ACCEPTANCE TESTING

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	2	1	0	1	4
Duplicate	1	0	0	0	1
External	0	3	0	2	5
Fixed	5	1	6	8	20
Not Reproduced	0	1	1	0	2
Skipped	0	1	0	2	3
Won't Fix	0	7	0	2	9
Totals	8	14	7	15	44

CHAPTER 9
RESULTS

PERFORMANCE METRICS

```
'Ambassador Grand 1800', 'Audi A4 1.8', 'Audi A4 2.0',
'Audi A4 3.0', 'Audi A4 30', 'Audi A4 35', 'Audi A4 New',
'Audi A5 Sportback', 'Audi A6 2.0', 'Audi A6 2.7', 'Audi A6 2.8',
'Audi A8 4.2', '...
                                                                array(['CNG', 'Diesel', 'Electric', 'LPG', 'Petrol'], dtype=object),
                                                                array(['Dealer', 'Individual', 'Trustmark Dealer'], dtype=object),
                                                                array(['Automatic', 'Manual'], dtype=object),
                                                                array(['First Owner', 'Fourth & Above Owner', 'Second Owner',
'Test Drive Car', 'Third Owner'], dtype=object)]),
                                          ['car_models', 'company_name',
                                           'fuel', 'seller_type',
                                           'transmission',
                                           'owner'])])),
        ('linearregression', LinearRegression())])
```

[26]:
```
y_pred=pipe.predict(X_Test)
```

[27]:
```
r2_score(Y_Test,y_pred)
```

t[27]:
```
0.8949529491096558
```

[28]:
```
import pickle
```

[29]:
```
pickle.dump(pipe,open('CarLinearRegressionModel.pkl','wb'))
```

[30]:
```
car.head
```

t[30]:

[31]:
```
car.head
```

CHAPTER 10

ADVANTAGES

- Very Fast prediction .
- Simple UI, just click the link and predict .
- Better accuracy without overfitting .
- High availability .

DISADVANTAGES

- If the actual value is higher then, the data points may have higher deviation in the `predicted value.
- User entered data are not stored in database .

CONCLUSION

Thus, by using Random Forest regressor, the model can predict the resale value of the car with maximum accuracy without overfitting. It also consumes low memory and faster than other regressors for this dataset. Also, deploying the model in IBM cloud allows us to use the model from the hosted website. In depth analysis and powerful computers can make this model more accurate.

FUTURE SCOPE

This project will be more useful in future, as renting and reselling of a car is becoming more common, there are some services such as cars24.com where we can buy and sell second hand cars, the customers of these services may need to estimate or predict the resale value of the cars based on its configuration.

APPENDIX

Source Code :
https://github.com/IBM-EPBL/IBM-Project-35390-1660284416/tree/main/Final%20Deliverables/PROJECT
GitHub:
https://github.com/IBM-EPBL/IBM-Project-35390-1660284416
Project Demo Link:
https://youtu.be/yHxr-zQD0ds