

# Car resale value prediction



**TEAM ID**

**PNT2022TMID04253**

**TEAM MEMBERS**

**SUBRAMANIAN K**

**KRISHNA KUMAR C**

**NIVETHA K**

**MAADHINI B**

# **INDEX**

## **1 INTRODUCTION**

1.1 Project Overview

1.2 Purpose

## **1) LITERATURE SURVEY**

2.1 Existing problem

2.2 References

2.3 Problem Statement Definition

## **3) IDEATION & PROPOSED SOLUTION**

3.1 Empathy Map Canvas

3.2 Ideation & Brainstorming

3.3 Proposed Solution

3.4 Problem Solution fit

## **4) REQUIREMENT ANALYSIS**

4.1 Functional requirement

4.2 Non-Functional requirements

## **5) PROJECT DESIGN**

5.1 Data Flow Diagrams

5.2 Solution & Technical Architecture

5.3 User Stories

## **6) PROJECT PLANNING & SCHEDULING**

6.1 Sprint Planning & Estimation Delivery Schedule

6.2 Sprint

6.3 Reports from JIRA

## 7) CODING & SOLUTIONING

(Explain the features added in the project along with code)

7.1 Feature 1

7.2 Feature 2 7.3 Database Schema (if Applicable)

## 8) TESTING

8.1 Test Cases

8.2 User Acceptance Testing

## 9) RESULTS

9.1 Performance Metrics

## 10) ADVANTAGES & DISADVANTAGES

## 11) CONCLUSION

## 12) FUTURE SCOPE

## 13) APPENDIX

Source Code

GitHub & Project Demo Link

# 1) INTRODUCTION TO PROJECT

---

## 1.1 )Project Overview:

Predicting the resale value of a car is not a simple task. It is trite knowledge that the value of used cars depends on a number of factors. The most important ones are usually the age of the car, its make (and model), the origin of the car (the original country of the manufacturer), its mileage (the number of kilometers it has run) and its horsepower. Due to rising fuel prices, fuel economy is also of prime importance. Unfortunately, in practice, most people do not know exactly how much fuel their car consumes for each km driven. Other factors such as the type of fuel it uses, the interior style, the braking system, acceleration, the volume of its cylinders (measured in cc), safety index, its size, number of doors, paint color, weight of the car, consumer reviews, prestigious awards won by the car manufacturer, its physical state, whether it is a sports car, whether it has cruise control, whether it is automatic or manual transmission, whether it belonged to an individual or a company and other options such as air conditioner, sound system, power steering, cosmic wheels, GPS navigator all may influence the price as well. Some special factors which buyers attach importance in Mauritius is the local of previous owners, whether the car had been involved in serious accidents and whether it is a lady-driven car. The look and feel of the car certainly contributes a lot to the price. As we can see, the price depends on a large number of factors.

## 1.2 )Purpose:

Predicting the price of used cars in both an important and interesting problem.

According to data obtained from the National Transport Authority, the number of cars registered between 2013 and 2020 has witnessed a spectacular increase of 234%. From 168, 524 cars registered in 2013, this number has now reached 35,07,014. With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. It is reported in that the sales of new cars has registered a decrease of 8% in 2020. In many developed countries, it is common to lease a car rather than buying it outright. A lease is a binding contract between a buyer and a seller (or a third party – usually a bank, insurance firm or other financial institutions) in which the buyer must pay fixed

installments for a pre-defined number of months/years to the seller/financer.

After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, it is of commercial interest to seller/financers to be able to predict the salvage value (residual value) of cars with accuracy. If the residual value is under-estimated by the seller/financer at the beginning, the installments will be higher for the clients who will certainly then opt for another seller/financer. If the residual value is over-estimated, the installments will be lower for the clients but then the seller/financer may have much difficulty at selling these high-priced used cars at this over-estimated residual value. Thus, we can see that estimating the price of used cars is of very high commercial importance as well. Manufacturers' from Germany made a loss of 1 billion Euros in their USA market because of miscalculating the residual value of leased cars. Most individuals in Mauritius who buy new cars are also very apprehensive about the resale value of their cars after certain number of years when they will possibly sell it in the used cars market.

## 2) LITERATURE SURVEY

---

### 2.1) EXISTING PROBLEMS:

The prices of new cars in the industry are 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. 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 offers this service, their prediction method may not be the best. Besides, different models and systems may contribute on predicting power for a used car's actual market value. It is important to know their actual market value while both buying and selling.

### Related work

Surprisingly, work on estimated the price of used cars is very recent but also very sparse. In her MSc thesis [3], Listiani showed that the regression mode build using support vector machines (SVM) can estimate the residual price of leased cars with higher accuracy than simple multiple regression or multivariate regression. SVM is better able to deal with very high dimensional data (number of features used to predict the price) and can avoid both over-fitting and under-fitting. In particular, she used a genetic algorithm to find the optimal parameters for SVM in less time. The only drawback of this study is that the improvement of SVM regression over simple regression was not expressed in simple measures like mean deviation or variance. In another university thesis, Richardson working on the hypothesis that car manufacturers are more willing to produce vehicles which do not depreciate rapidly. In particular, by using a multiple regression analysis, he showed that hybrid cars (cars which use two different power sources to propel the car, i.e. they have both an internal combustion engine and an electric motor) are more able to keep their value than traditional vehicles. This is likely due to more environmental concerns about the climate and because of its higher fuel efficiency. The importance of other factors like age, mileage, make and MPG (miles per gallon) were also considered in this study. He collected all his data from various websites.

## 2.2 REFERENCES:

### **1. CAR PRICE PREDICTION [Abhay Yadav, Chavi Ralhan ET AL, 2022]**

India has a considerable size of car sales on top of the world day-to-day many buyers usually sell their cars after using for the time to another buyer, they name them as second possessor. Numerous platforms such as cars24.com, OLX.com that come up with these buyers with a platform where they can sell their old cars, but what should be the price of the car, this is the long-lasting query ever by using Machine Learning algorithms and they lead a response to this issue. Using a history of previous used car sales data and machine learning methodologies like Supervised Learning, they used to predict a fair price for the car. They also used machine learning techniques like Random Forest and Extra Tree Regression

### **2. USED CAR PRICE PREDICTION AND LIFE SPAN [Aditya Nikhade, Rohan Borde, 2021]**

The predictions are based on dataset collected from various websites and Kaggle Websites mostly. This project will compare all this data to all regression algorithms and performance of various machine learning algorithms such as Linear Regression, Ridge Regression, Decision tree Regressor and choose the best out of it. Depending on various parameters the project will determine the price of a car and compare the prices of old cars with new cars. The lifespan of the car can be determined using Government regulations and Company claims. Apart from various factors, they also consider GPS navigator to predict the price of the car.

### **3. Car Price Prediction Using Machine Learning [Ketan Agrahari, Ayush Chaubey ET AL, 2021]**

The rise of online websites and other tools like it have made it easier for both buyers and sellers to get a better understanding of the factors that determine the market value of a used car. Based on a set of factors, Machine Learning algorithms may be used to forecast the price of any automobile. The cost is calculated using the amount of characteristics. They used linear regression and lasso regression to develop a price model for used automobiles in a comparative research. The main goal of this study is to discover the best predictive model for estimating the price of a used car.

### **4. Used Car Price Prediction using K-Nearest Neighbor Based Model [Samruddhi, Ashok Kumar, 2020]**

In this paper, they proposed a model to estimate the cost of the used cars using the K nearest neighbor algorithm which is simple and suitable

for small data set. Here, they have collected a used cars dataset and analyzed the same. The data was trained by the model and examined the accuracy of the model among different ratios of trained and test set. The same model is cross-validated for assessing the performance of the model using the K- Fold method which is easy to understand and implement. They have used the K nearest Neighbor algorithm and got accuracy 85% where the accuracy of linear regression is 71%. The proposed model is also validated with 5 and 10 folds by using K Fold Method. The experimental analysis shows that the proposed model is fitted as the optimized model

##### **5. Car Price Prediction using Machine Learning Techniques [Enis Gegic, Becir Isakovic ET AL, 2019]**

The major step in this prediction process is collection and preprocessing of the data. In this research ,PHP scripts were built to normalize, standardize and clean data to avoid unnecessary noise for machine learning algorithms. To build a model that predicting the price of used cars in Bosnia and Herzegovina, they applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal autopijaca.ba using a web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set.

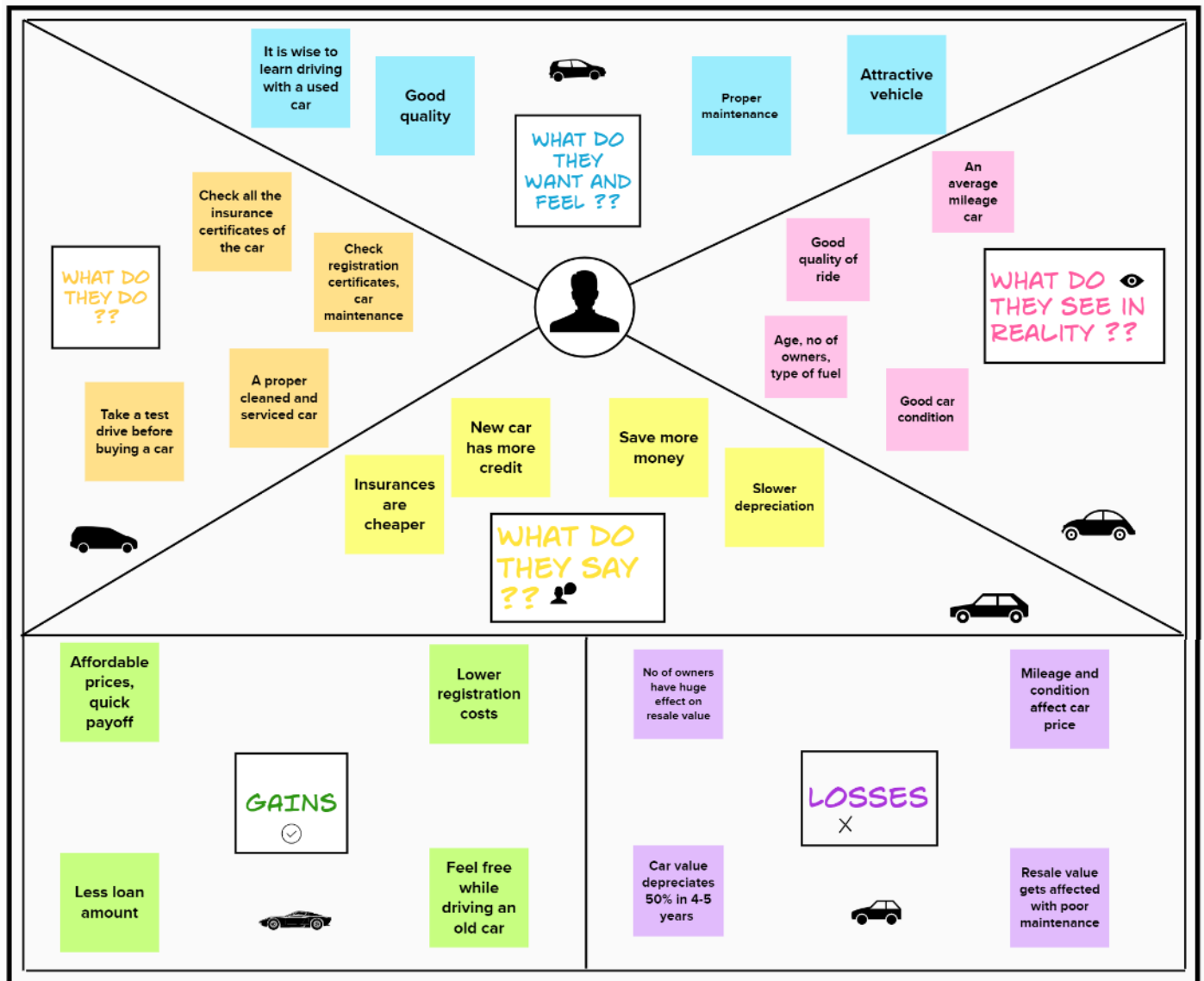
##### **PROBLEM STATEMENT DEFINITION:**

<b>Problem Statement (PS)</b>	<b>I am (Customer)</b>	<b>I'm trying to</b>	<b>But</b>	<b>Because</b>	<b>Which makes me feel</b>
PS-1	Businessman	Buy a car for my personal use	Don't have time to search and predict	Of lot of work pressure	Frustrated
PS-2	Travel agent owner	Buy a car for my travel agency	Can't predict the value	Changing market value	Depressed
PS-3	Family man	Buy a car for my family use	Not able to choose the right one	Trying to find a Car that fits my budget	Sorrowful
PS-4	Dealer	Sell a car to the customers	Not able to predict the exact value	Different car has different features	Confused
PS-5	Racer	Buy a sports car	The cars are costly	Higher horse power	Distressed



### 3) IDEATION AND PROPOSED SOLUTION

#### 3.1) EMPATHY MAP CANVAS:



## 3.2) IDEATION AND BRAINSTORMING:



### Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

🕒 10 minutes to prepare

🕒 1 hour to collaborate

👤 2-8 people recommended



#### Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

🕒 10 minutes

A

#### Team gathering

Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

B

#### Set the goal

Think about the problem you'll be focusing on solving in the brainstorming session.

C

#### Learn how to use the facilitation tools

Use the Facilitation Superpowers to run a happy and productive session.

[Open article](#)



1

#### Define your problem statement

•In this project we have used different algorithms with different techniques for developing Car resale value prediction systems considering different features of the car.

🕒 5 minutes

#### PROBLEM

Analyse and predict the resale value of cars using different algorithms



#### Key rules of brainstorming

To run a smooth and productive session



Stay in topic.



Encourage wild ideas.



Defer judgment.



Listen to others.



Go for volume.



If possible, be visual.

2

## Brainstorm

Write down any ideas that come to mind that address your problem statement.

🕒 10 minutes

### Subramanian K

Make high quality dataset  
 Verify insights from previous models  
 Various vehicular conditions  
 Increasing efficiency by various approach

### Krishna Kumar C

The proposed model should have high accuracy  
 Using machine learning to utilise data to predict more accurately  
 This method works well if you have a common car with common features  
 The nature of ML is to train past data to predict unseen data

### Nivetha K

Using ML will be easier to predict resale value  
 Scattered plot chart is used for better understanding  
 Data cleaning methods are used  
 User interface is used to display price of car

### Maadhini B

Using Forest regression algorithm  
 Using time series analysis and prediction  
 By finding car engine condition  
 Using flask application

3

## Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you can break it up into smaller sub-groups.

🕒 20 minutes

### Grouping based on dataset

Use large amount of dataset  
 Data cleaning methods are used  
 Verify insights from previous models

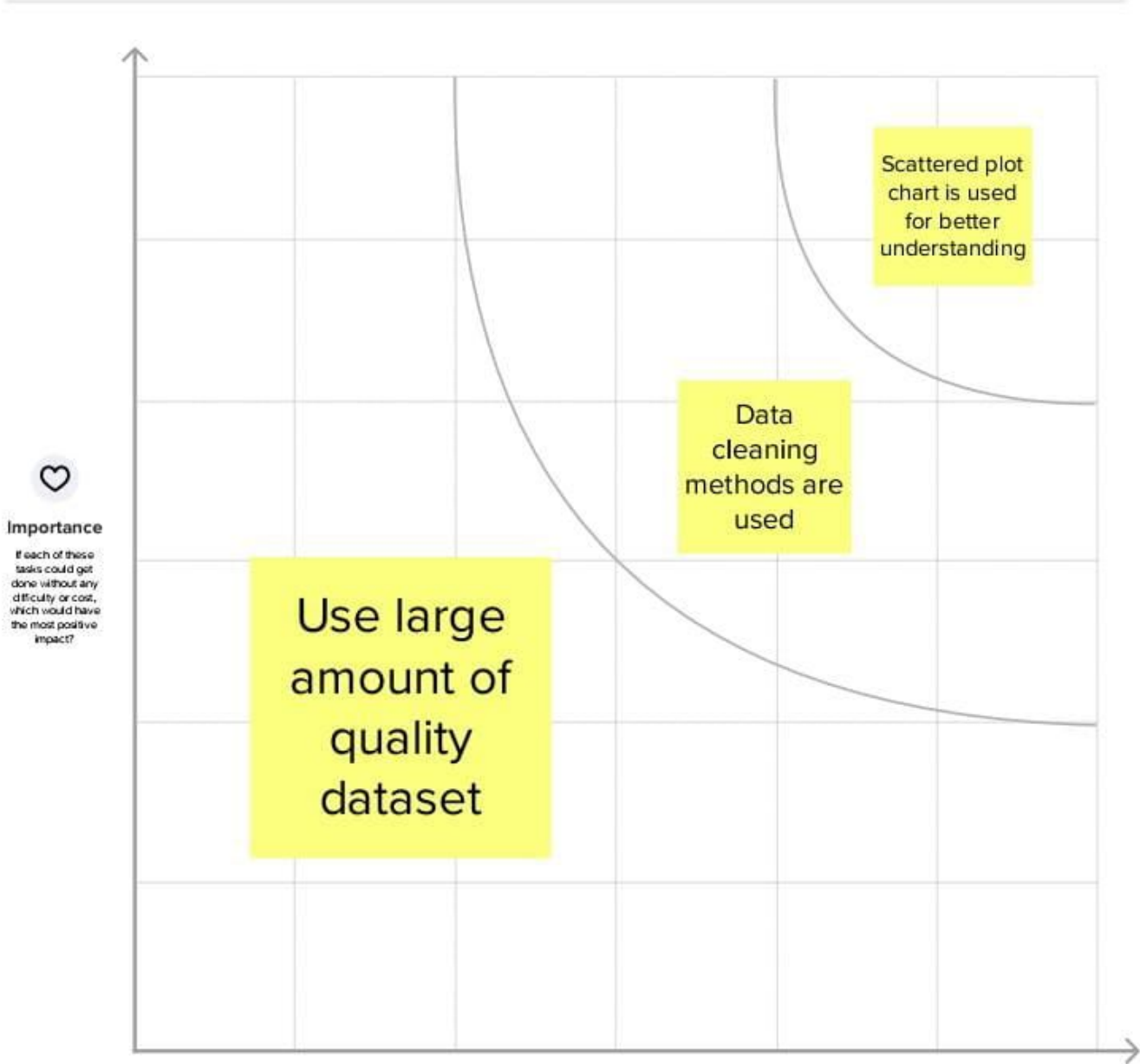
### Grouping based on model

Scattered plot chart is used for better understanding  
 Using Forest regression algorithm  
 Using time series analysis  
 Using flask application

# Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

🕒 20 minutes





## After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

---

### Quick add-ons



#### Share the mural

Share a **view link** to the mural with stakeholders to keep them in the loop about the outcomes of the session.



#### Export the mural

Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

---

### Keep moving forward



#### Strategy blueprint

Define the components of a new idea or strategy.

[Open the template →](#)



#### Customer experience journey map

Understand customer needs, motivations, and obstacles for an experience.

[Open the template →](#)



#### Strengths, weaknesses, opportunities & threats

Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.

[Open the template →](#)

---

### 3.3)PROPOSED SOLUTION:

#### Proposed Solution Template:

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To Predict the value of the used car using Data Science
2.	Idea / Solution description	Pre-owned car sale is more popular in developing country. People Who plan to purchase used car often struggle to find a within a budget as well as to predict the price of the second-hand car. So our project helps a potential buyer to estimate the price of a Second-hand car. Analysis Data using various Machine learning Algorithms.
3.	Novelty / Uniqueness	We predict Used Car price mainly based on the car condition and mitigation level of quality, capitalization chart is provided accordingly.
4.	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.
5.	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.
6.	Scalability of the Solution	Whatever may be the vehicle type or count of vehicle, this system predicts the appropriate resale value If Multiple user access the system at same time, it process scalable.

## 3.4) PROBLEM SOLUTION FIT:

Project Title: Car Resale Value Prediction

Project Design Phase-I - Solution Fit Template

Team ID: PNT2022TMID04253

Define CS, fit into CC	<b>1. CUSTOMER SEGMENT(S)</b> <span>CS</span> Who is your customer? i.e. working parents of 0-5 y.o. kids  Used car sellers	<b>6. CUSTOMER CONSTRAINTS</b> <span>CC</span> What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.  <ul style="list-style-type: none"> <li>To determine the worthiness of the car by their own within few minutes</li> <li>A loss function is to be optimized by spending money for dealers, brokers to buy or sell a car.</li> </ul>	<b>5. AVAILABLE SOLUTIONS</b> <span>AS</span> Which solutions are available to the customers when they face the problem?  or What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking need to get the job done? What have they tried in the past?  <ul style="list-style-type: none"> <li>In the past User cannot find the value of used car buy their own without prior knowledge about cars.</li> <li>A person who don't know much about the car can also make predictions for used cars easily.</li> </ul>	Explore AS, differentiate
	<b>2. JOBS-TO-BE-DONE / PROBLEMS</b> <span>J&amp;P</span> Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.  To build a supervised machine learning model using regression algorithms for forecasting the value of a vehicle based on multiple attributes such as  <ul style="list-style-type: none"> <li>Condition of Engine</li> <li>Year of Registration</li> <li>Kilometers</li> <li>Number of Owner</li> </ul>	<b>9. PROBLEM ROOT CAUSE</b> <span>RC</span> What is the real reason that this problem exists? What is the back story behind the need to do this job? i.e. customers have to do it because of the change in regulations.  <ul style="list-style-type: none"> <li>The price predicted by the dealers or brokers for used car is not trustful</li> <li>Users can predict the correct valuation of the car remotely without human intervention like car dealers.</li> <li>User can eliminate the valuation predicted by the dealer</li> </ul>	<b>7. BEHAVIOUR</b> <span>BE</span> What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)  <ul style="list-style-type: none"> <li>The History of Your Car's condition and documents produced by them will be Suspicious.</li> <li>The model is to be built would give the nearest value of the vehicle by eliminating anonymous value predicted by using humans.</li> </ul>	
<b>3. TRIGGERS</b> <span>TR</span> What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.  Users can predict the correct valuation of the car by their own like Olxcars, Cars24 and other car resale value prediction websites by using model, year, owner, etc.	<b>10. YOUR SOLUTION</b> <span>SI</span> If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality. If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.  <ul style="list-style-type: none"> <li>The main aim of this project is to predict the price of used cars using the Machine Learning (ML) algorithms and collection data's about different cars.</li> </ul>	<b>8. CHANNELS of BEHAVIOUR</b> <span>CH</span> <b>8.1 ONLINE</b> What kind of actions do customers take online? Extract online channels from #7  <b>8.2 OFFLINE</b> What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.  <ul style="list-style-type: none"> <li>Customer should predict the worth of the car by using different parameters given by the owner.</li> </ul>		
Identify strong TR & EM	<b>4. EMOTIONS: BEFORE / AFTER</b> <span>EM</span> How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure > confident, in control - use it in your communication strategy & design.  Before: <ul style="list-style-type: none"> <li>User will be in fear about the biased values predicted by the humans based on the condition of the car.</li> </ul> After: <ul style="list-style-type: none"> <li>User can determine the worthiness of the car by their own without human intervention.</li> </ul>	The project should take parameters related to used car as inputs and enable the customers to make decisions by their own.	<ul style="list-style-type: none"> <li>User Should confirm the details provided about the vehicle in RTO online.</li> <li>User can decide by seeing the exterior and interior condition of the car.</li> <li>User can test the performance of the car and to buy it up in a affordable price based on its condition.</li> </ul>	

## 4)REQUIREMENT ANALYSIS

---

### 4.1)FUNCTIONAL REQUIREMENTS:

- 1) FR-1 **Car's Data** Fill the required data.
- 2) FR-2 **Prediction** Analyse the car's price.
- 3) FR-3 **Notify** Provide the analysed data via notification.

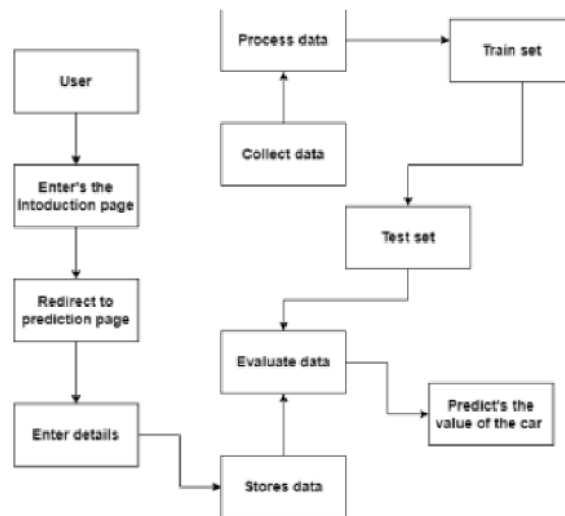
### 4.2)NON-FUNCTIONAL REQUIREMENTS:

- 1) NFR-1 **Usability** Used to predict car's price.
- 2) NFR-2 **Reliability** Prediction of accurate car's resale value.
- 3) NFR-3 **Performance** Reducing overall load time.
- 4) NFR-4 **Availability** Can be accessed anytime and anywhere.  
Available for everyone.
- 5) NFR-5 **Scalability** Multiple users can access the website at same time.



## 5) PROJECT DESIGN

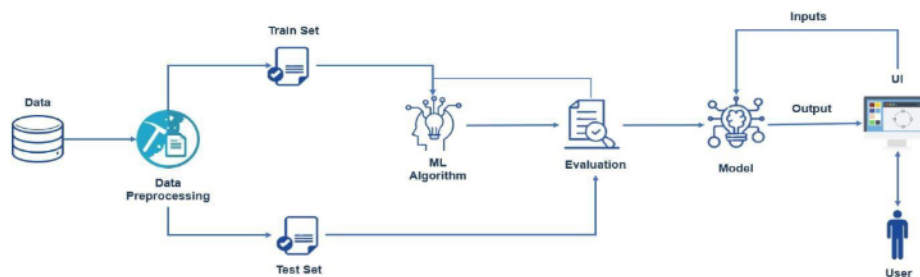
### 5.1) DATA FLOW DIAGRAM:



### 5.2) SOLUTION AND TECHNICAL ARCHITECTURE:

#### TECHNICAL ARCHITECTURE:

The Deliverable shall include the architectural diagram as below and the information as per the table 1 & table 2



### 5.3)USER STORIES:

#### **User Stories**

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

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through mobile number	I can register & access the dashboard using mobile number	Medium	Sprint-1
		USN-4	As a user, I can register for the application through Gmail	I can register & access the dashboard	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	I can access the dashboard	High	Sprint-1
	Dashboard	USN-6	As a user, I can visit the home page	I can edit my profile	Medium	Sprint-1
	Car details	USN-7	As a user, I should give the car details like car model, engine and fuel type, etc...	I should the give car details	High	Sprint-2
	Car Price	USN-8	As a user, I can view the current rate of the used car price	I know the car price	High	Sprint-4
Admin	Model Building	USN-9	As an admin, I should train and test the given dataset	I can build the train and test model	High	Sprint-3
	Prediction Chart	USN-10	As an admin, I should send the prediction chart and the nearest service centre location	I can provide chart	Medium	Sprint-3
	Predict	USN-11	As an admin, I will predict the price	I can predict the car price	High	Sprint-4

## 6.PROJECT PLANNING AND SCHEDULING

---

### 6.1)SPRINT PLANNING AND ESTIMATION:

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring the, technical papers, research publications etc	29 October 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	30 October 2022
Ideation	List the by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	30 October 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	4 October 2022
Problem Solution Fit	Prepare problem - solution fit document.	12 October 2022
Solution Architecture	Prepare solution architecture document.	12 October 2022
Customer Journey	Prepare the customer journey maps to understand the user interactions & experiences with the application (entry to exit).	5 November 2022
Functional Requirement	Prepare the functional requirement document	5 November 2022
Data Flow Diagrams	Draw the data flow diagrams and submit for review.	12 October 2022
Technology Architecture	Prepare the technology architecture diagram	5 November 2022
Prepare Milestone & Activity List	Prepare the milestones & activity list of the project.	16 November 2022
Project Development - Delivery of Sprint-1, 2, 3 & 4 Develop & submit the d	Develop & submit the developed code by testing it.	IN PROGRESS...

## 6.2)SPRINT DELIVERY SCHEDULE:

### ProductBacklog, Sprint Schedule, andEstimation (4Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement(Epic)	UserStory Number	UserStory/Task	Story Points	Priority	Team Members
Sprint-1	Dataset reading and Pre processing	USN-1	Cleaning the dataset and splitting to dependent and independent variables	2	High	NIVETHA K
Sprint-2	Building the model	USN-2	Choosing the appropriate model for building and saving the model as pickle file	1	High	NIVETHA K
Sprint-3	Application building	USN-3	Using flask deploying t..e ML model	2	Medium	MAADHINI B
Sprint-4	Train the model in IBM	USN-4	Finally train the model on IBM cloud and deploy the application	2	Medium	MAADHINI B

### ProjectTracker, Velocity & Burndown Chart: (4Marks)

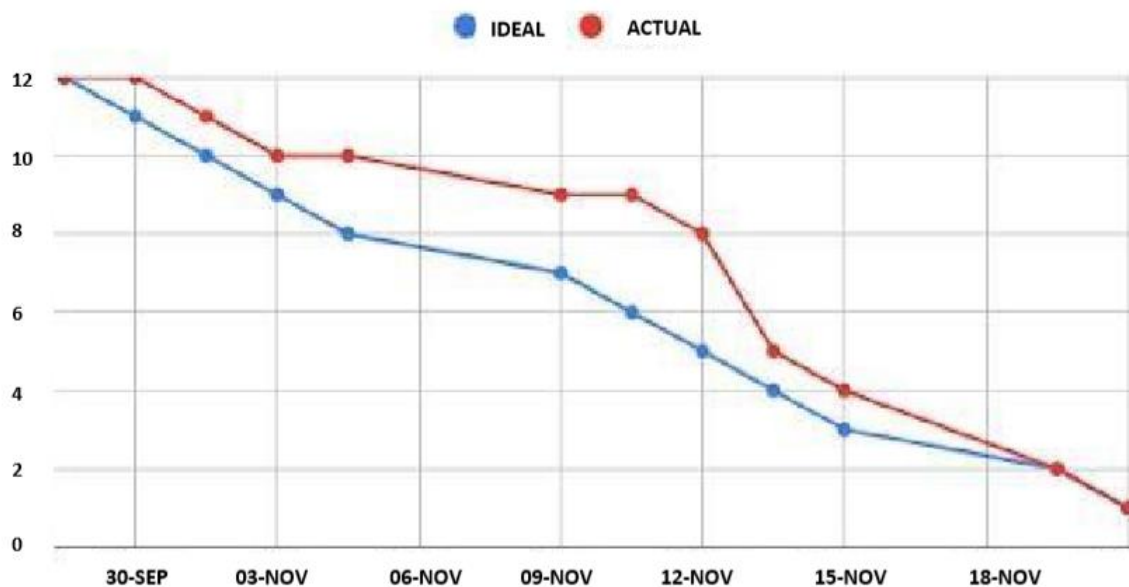
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed(as on PlannedEndDate)	SprintReleaseDate (Actual)
Sprint-1	15	5 Days	24Oct2022	29Oct2022	15	29 Oct 2022
Sprint-2	15	5 Days	31 Oct 2022	05Nov 2022	15	05 Nov 2022
Sprint-3	15	5 Days	07 Nov 2022	12 Nov 2022	15	12 Nov 2022
Sprint-4	15	5 Days	14 Nov 2022	19 Nov 2022	15	18 Nov 2022

#### Velocity:

We have a 5-day sprint duration, and the velocity of the team is 15 (points per sprint). The team's average velocity (AV) per iteration unit (story points per day)

$$\text{Actual Velocity} = \frac{\text{Sprint Duration}}{\text{Velocity}} = \frac{15}{5} = 3$$

## 6.3)REPORTS FROM JIRA SOFTWARE:



# 7) CODING AND SOLUTIONING

## 7.1)FEATURE:

### APPLYING XG-BOOST FOR DATASET.

```
In [2]: ##import libraries
import pandas as pd
import numpy as np
import matplotlib as plt
from sklearn.preprocessing import LabelEncoder
import pickle
import seaborn as sns
```

```
In [3]: car=pd.read_csv(r"C:\Users\narma\OneDrive\Desktop\Project\autos.csv",header=0,sep=',',encoding='Latin1',)
```

```
In [4]: car.sample()
```

```
Out[4]: Unnamed: 0 car_name brand model min_cost_price max_cost_price vehicle_age km_driven seller_type fuel_type transmission_type mileage engine max_power
```

4992	6333	Maruti Swift Dzire	Maruti Swift Dzire	688000.0	1024000.0	3	18000	Dealer	Petrol	Automatic	21.21	1197	81.8
------	------	--------------------	--------------------	----------	-----------	---	-------	--------	--------	-----------	-------	------	------

```
In [26]: car.info()
```

```
RangeIndex: 15411 entries, 0 to 15410
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  --
0   brand            15411 non-null  object
1   model            15411 non-null  object
2   vehicle_age      15411 non-null  int64
3   km_driven        15411 non-null  int64
4   fuel_type        15411 non-null  object
5   transmission_type 15411 non-null  object
```

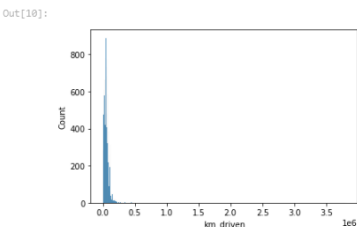
```
Out[6]: brand model vehicle_age km_driven fuel_type transmission_type mileage engine max_power seats Owner_type selling_price
12402  Maruti Swift 12 54000 Petrol Manual 16.1 1298 88.2 5 1 360000
```

```
In [8]: car1=car.copy()
car1=car1.drop_duplicates(['brand','model','vehicle_age','km_driven','fuel_type','transmission_type','mileage','engine','max_power','seats','Owner_type'])
```

```
In [9]: car1.shape
```

```
Out[9]: (15331, 12)
```

```
In [10]: sns.histplot(data=car1,x='km_driven')
```



```
In [107]: #xgboost
from xgboost import XGBRegressor
xgbr=XGBRegressor()
xgbr.fit(X_train,Y_train)
```

```
Out[107]: XGBRegressor(base_score=0.5, booster='gbtree', callbacks=None,
colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1,
early_stopping_rounds=None, enable_categorical=False,
eval_metric=None, gamma=0, gpu_id=-1, grow_policy='depthwise',
importance_type=None, interaction_constraints='',
learning_rate=0.300000012, max_bin=256, max_cat_to_onehot=4,
max_delta_step=0, max_depth=6, max_leaves=0, min_child_weight=1,
missingnan, monotone_constraints='()', n_estimators=100, n_jobs=0,
num_parallel_tree=1, predictor='auto', random_state=0, reg_alpha=0,
reg_lambda=1, ...)
```

```
In [108]: xgbr.score(X_train,Y_train)
```

```
Out[108]: 0.951034958541807
```

```
In [115]: xgbr.score(X_test,Y_test)
```

```
Out[115]: 0.8955792472380257
```

```
In [116]: #saving the model
import pickle
pickle.dump(xgbr,open("boost.pkl","wb"))
```

## 8) TESTING

### 8.1)TESTCASES:

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status
elcome_TC_001	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_002	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_003	Functional/Result	Notification	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

#### Test Scenarios

- 1 Verify redirecting to flask application it must appear before the main page.
- 2 Verify the UI elements in main page
- 3 Verify user is able to get the value for the given details.

#### Search

- 1 Verify user is able to search elements in dropdown.
- 2 Verify user is able to see suggestions based on numbers entered in search box
- 3 Verify user is able to see related auto suggestions displaying based on keyword entered in search box
- 4 Verify user is able to see search detailed page when nothing entered in textbox

## 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 Car Resale Value Prediction 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	Severity1	Severity2	Severity3	Severity4	Subtotal
By Design	10	2	1	0	13
Duplicate	2	0	0	0	2
External	1	3	0	0	4
Fixed	13	2	1	2	18
Not Reproduced	1	0	0	0	1
Skipped	0	0	1	0	1
Won't Fix	5	2	1	0	8
Totals	31	9	4	2	47

### 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	5	0	0	5
Client Application	10	0	0	10
Security	1	0	0	1

Exception Reporting	4	0	0	4
Final Report Output	5	0	0	5
Version Control	2	0	0	2

## 9) RESULTS

---

### 9.1) PERFORMANCE METRICES (IN PERCENTAGE):

- |    |                                |   |      |
|----|--------------------------------|---|------|
| 1) | Linear Regression score(train) | - | 74   |
| 2) | Linear Regression score(test)  | - | 74.8 |

- |    |                                       |   |      |
|----|---------------------------------------|---|------|
| 1. | Random Forest Regression score(train) | - | 98.2 |
| 2. | Random Forest Regression score(test)  | - | 88   |

- |    |                       |   |      |
|----|-----------------------|---|------|
| 1. | XG-Boost score(train) | - | 95   |
| 2. | XG-Boost score(test)  | - | 88.5 |



## 10) ADVANTAGES AND DISADVANTAGES

---

**Used car sellers (dealers):** They are one of the biggest target group that can be interested in results of this study. If used car sellers better understand what makes a car desirable, what the important features are for a used car, then they may consider this knowledge and offer a better service.

**Online pricing services:** There are websites that offers an estimate value of a car. They may have a good prediction model. However, having a second model may help them to give a better prediction to their users. Therefore, the model developed in this study may help online web services that tells a used car's market value.

**Individuals:** There are lots of individuals who are interested in the used car market at some points in their life because they wanted to sell their car or buy a used car. In this process, it's a big corner to pay too much or sell less than it's market value.

### **Advantages:**

- 1) Able to give accurate and acceptable price for both buyer and seller.
- 2) Have range of option on buying on budget.
- 3) Helps in saving money than giving to brokerage.

### **Disadvantages:**

- 1) Poor checking and invalid information affect the value of prediction.
- 2) Cars are limited usage vehicles some people only could afford this basis on knowledge based purchasing.

## 11) CONCLUSION

---

- In the given guided project I understood the problem to classify if it is a regression or a classification kind of problem.
- I also came to know how to pre-process the data using different data-preprocessing techniques
- Not only this, I also grasp the knowledge about applying different algorithms according to the dataset
- I also learn about the features of flask application

## 12)FUTURE SCOPE

---

- With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. In many developed countries, it is common to lease a car rather than buying it outright. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, it is of commercial interest to sellers/financers to be able to predict the salvage value (residual value) of cars with accuracy.
- In order to predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale value of the vehicle.

## 13) APPENDIX

---

### GITHUB LINK:

<https://github.com/IBM-EPBL/IBM-Project-26323-1660025101>

### PROJECT DEMO LINK:

<https://drive.google.com/drive/u/0/folders/1j9i3TeJkVCF3Exc0hByHSTaEmb6UVVsJ>