CAR RESALE VALUE PREDICTION

NALAIYA THIRAN PROJECT BASED LEARNING

ON

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

Submitted By

TEAM ID: PNT2022TMID31751

HARIPRIYA R	**711719205020**
JEEVA H	**711719205022**
KUBERA PRIJESH D	**711719205028**
SNEHA N S	**711719205052**

in partial fulfilment for the award of the degree

of

Bachelor of Technology

in

Information Technology

KGiSL Institute Of Technology

II

Coimbatore - 641 035

Affiliated to Anna University, Chennai - 600 025.

i

KGiSL Institute Of Technology

Coimbatore - 641 035

Affiliated to Anna University, Chennai - 600 025

INTERNAL MENTOR

Assistant Prof Suresh Kumar C

INDUSTRY MENTOR

Prof Sweta

ABSTRACT

Due to the unprecedented number of cars being purchased and sold, used car price prediction is a topic of high interest. Because of the affordability of used cars in developing countries, people tend more purchase used cars. A primary objective of this project is to estimate used car prices by using attributes that are highly correlated with a label (Price). To accomplish this, data mining technology has been employed. Null, redundant, and missing values were removed from the dataset during pre-processing. In this supervised learning study, three regressors (Random Forest Regressor, Linear Regression, and Bagging Regressor) have been trained, tested, and compared against a benchmark dataset. Among all the experiments, the Random Forest Regressor had the highest score at 95%, followed by 0.025 MSE, 0.0008 MAE, and 0.0378 RMSE respectively. In addition to Random Forest Regression, Bagging Regression performed well with an 88% score, followed by Linear Regression having an 85% mark. A train-test split of 80/20 with 40 random states was used in all experiments. The researchers of this project anticipate that in the near future, the most sophisticated algorithm is used for making predictions, and then the model will be integrated into a mobile app or web page for the general public to use.

TABLE OF CONTENTS

1. **INTRODUCTION**

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. For this project, we developed the car resale value using a variety of algorithms and methodologies.prediction algorithms that take into account the car's many qualities. In a nutshell, automobile resale value prediction enables users to forecast the resale worth of a car based on a variety of factors, such as the number of miles traveled and the type of fuel used. The project focuses on getting input like kilometers a vehicle has run, year of manufacture, the owner, number of services the car has been given since its purchase, previous damages and so on.

1.1 PROJECT 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 kilometers driven, fuel type, etc.

1.2 PURPOSE

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.

This resale value prediction system is made for general purpose to just predict the amount that can be accurately acquired by the user. We try to predict the amount of resale by best 90% accuracy so the user can get an estimated value before he resales the car and doesn't make a deal in loss.

2. **LITERATURE SURVEY**

2.1 EXISTING PROBLEM

1.Car Price Prediction using Machine Learning

This proposed system helps to study car price detection and resale prediction. The results showed that there is a positive correlation between price and kilometers traveled, year of registration and kilometers traveled and a negative correlation between price and year of registration.

2.Used Car Price Prediction using K-Nearest Neighbor Based Model

This paper proposed a supervised machine learning model using KNN algorithm to analyze the price of used cars. We trained our model with data of used cars which is collected from the Kaggle website.

3.Prediction of Prices for Used Car by using Regression Models

This proposed approach aids in the investigation of a pricing evaluation model based on big data analysis.Given time itself could have a significant impact on the pricing of cars, the data was gathered over a shorter period of time than a month

4.Prediction of Used Car Price Based on Supervised Learning Algorithm (2021)

Extra Trees Regressor, Random Forest Regressor was used.The algorithm was optimized by using the hyper parameter function.

5.Car resale value prediction using quantify qualitative data and knowledge-based system

The difficult procedure has been the main method for acquiring knowledge.Explore different data types of car data.

6.Vehicle Price Prediction using SVM Techniques

The study sector has seen a rise in the popularity of car price forecasting, which calls for significant effort and knowledge of the subject matter specialists. This the merit as follows, distinct machine learning algorithms of varying performances.

7.Car resale value prediction using quantify qualitative data and knowledge-based system

In this paper which is titled Doan Van thai(2019) quantify qualitative data and knowledge-based systems. The primary technique for knowledge acquisition has Been the time-consuming process.

8.Used Car price prediction (2021)

The data in the data set are used to train the system, which then creates a model. The inputs are sent to the system, which is then tested to see how it functions.

2.2 REFERENCES

[1].M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the mirai botnet," in Proc. of USENIX Security Symposium, 2017.

[2].Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani,―Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization‖, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal,January 2018

[3].Hossein Hadian Jazi, Hugo Gonzalez, Natalia Stakhanova, and Ali A. Ghorbani. "Detecting HTTP-based Application Layer DoS attacks on Web Servers in the presence of sampling." Computer Networks, 2017

[4]. A. Shiravi, H. Shiravi, M. Tavallaee, A.A. Ghorbani, Toward developing a systematic approach to generate benchmark datasets for intrusion detection, Comput. Security 31 (3) (2012) 357–374.

[5].Z. He, T. Zhang, and R. B. Lee, ―Machine Learning Based DDoS Attack Detection from Source Side in Cloud,‖ in Proceedings of the 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud), pp. 114–120, New York, NY, USA, June 2017

[6].R. Doshi, N. Apthorpe and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, 2018, pp. 29-35.

[7].Jerome H. Friedman, (2002), Stochastic gradient boosting, Computational Statistics & Data Analysis, 38, (4), 367-378

[8].Friedman, Jerome H. Greedy function approximation: A gradient boosting machine. Ann. Statist. 29 (2001), no. 5, 1189--1232

## 2.3 PROBLEM STATEMENT DEFINITION

Customer Problem Statement

There is a critical need to close the gap between sellers and buyers due to the enormous demand for used cars and the shortage of professionals who can evaluate the proper valuation. This project focuses on developing a system that can impartially and accurately forecast the resale value of cars based on simple features like miles driven, the year of purchase, fuel type, etc. In this research, we have developed automotive resale value prediction systems using machine learning approaches while taking into account various car attributes. Only a few features are currently used to forecast a car's resale value. This can be expanded to include more features and input sets.

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	Entrepreneur	Buy a used car	I am unaware of the varieties available	I don't have correct guidance	Exhausted
PS-2	Normal man	Find a good second hand car	I am unaware about the price and other factors	I am unable to know the current updates	Stressed
PS-3	Owner of a car	Sell	I don't know about its current value	Of its ear and tear	Confused

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors and attitudes. It is a useful tool to help teams better understand their users. Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.
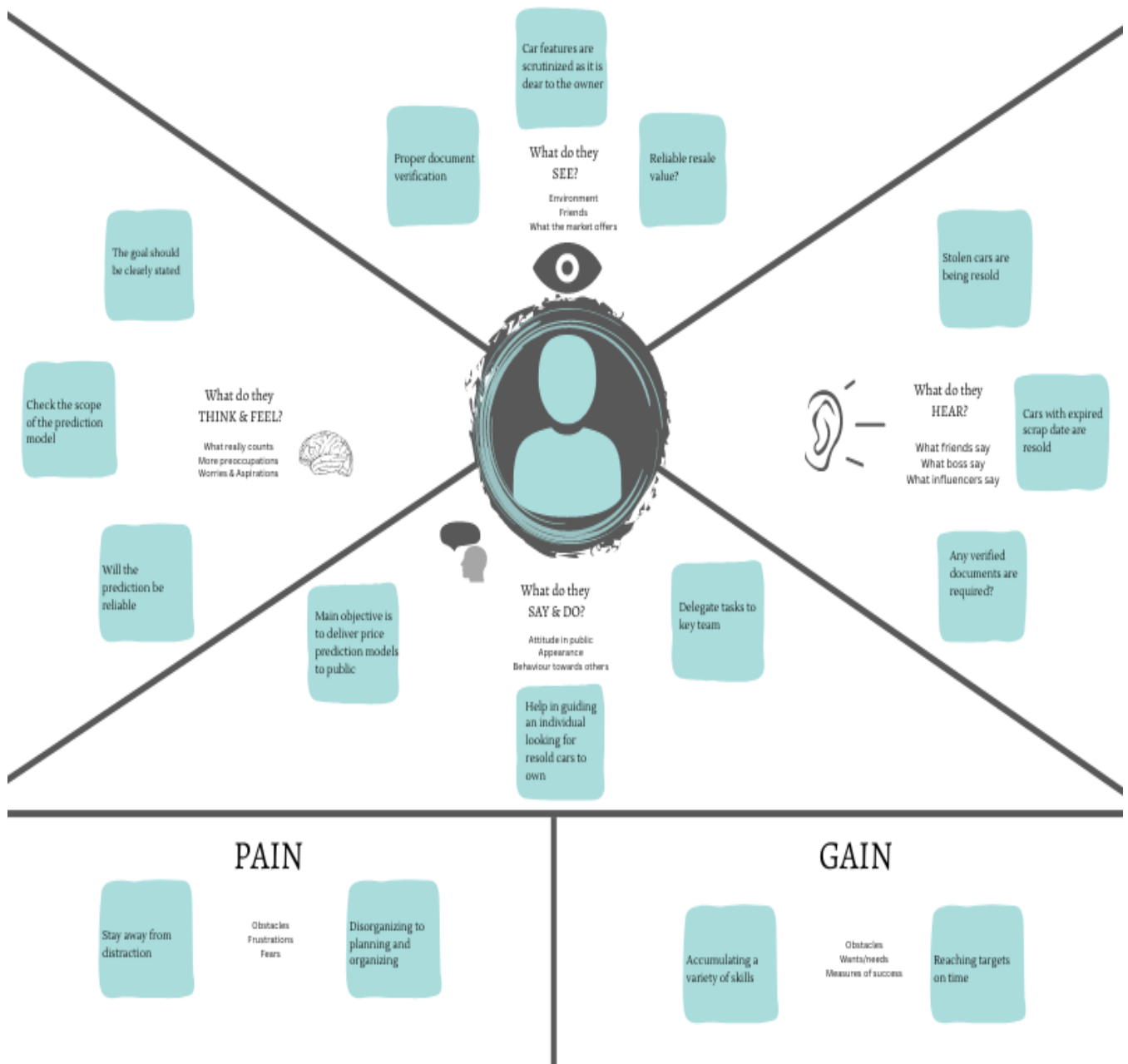
3.2 IDEATION & BRAINSTORMING

Step 1: Team Gathering, collaborating & selecting the problem statement

Step 2: Brainstorming, Idea Listing & Grouping

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 and break it up into smaller sub-groups.

🕐 20 minutes

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

3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Therefore, anyone who wants to sell their car must bring it to the appropriate corporate workshop or schedule an appointment with the business to acquire a price estimate. It takes a lot of time and energy to complete this process.
2.	Idea / Solution description	A used buy is more economical as well as more practical, especially for first-timers. A used automobile may easily accommodate practically all types of consumers, as long as you don't absolutely need the newest model on the market or just want the fragrance of a new car.
3.	Novelty / Uniqueness	For a higher car resale value, appearance matters. Good car resale value is guaranteed by a service. Keep all documents organised. Reselling cars is novel. Find Novelty Tata's Phone Numbers, Address, Photos, and Maps.
4.	Social Impact / Customer Satisfaction	Became obsessed with customer feedback, Create a sense of convenience, Deliver fast responses, satisfaction is a company –wide focus. Customer Satisfaction Look and Style Fuel consumption Pulling Power Seating Capacity Riding Comfort Safety Features Speed Shock Absorbs & transmission Tyre mileage Braking Efficiency
5.	Business Model (Revenue Model)	How to start a car merchant business. Generally, it is considered that if you want to start a car merchant business, you need a huge capital to invest. Dealer license.

		Location of the business. Keep a watch on the market. Make your catalog. Use a perfect marketing strategy.
6.	Scalability of the Solution	The size of the used car market in India was over 4.4 million units in 2020, according to Statista. The startup has managed to strive ahead by leveraging a robust managed marketplace business model, while proving that it is economically viable and independent of scale due to the use of technology, economy of scale, economy of scope, asset light, and network effects.

3.4 PROBLEM SOLUTION FIT

Define CS, fit into CC

1. CUSTOMER CS
- Common people
- Business Women
- Entrepreneur
- First time car buyer

6. CUSTOMER CONSTRAINTS —
Anxiety-customer began to get anxious when they still no idea about what they have found
Mysteries-they might Called it mysteries which they can't able to do.

5. AVAILABLE AS
- By searching in online websites.
- By gathering the information from the peoples and come to understanding.

Explore AS, differenti

2. JOBS-TO-BE-DONE / J&R
- Giving the necessary information for particular thing which needs for customer
- Solving customer doubts

9. PROBLEM ROOT R
- Lack of study in the sequence of things
- Unaware of the object
- New to environment

7. B
- Leased car need to be retained in good condition to avoid wear and tear penalties.
- Watch out for selling scams

Focus on J&P, tap into BE,

3. TRIGGERS TR
When it comes to motor vehicles, all the time people are posting pictures of the car as they do their Sunday drive or even just because it has had a wash. We have all seen the slamming cars get online when they break down! We trust these people to lead us to the right vehicle and to give us advice to help our buying decisions.

10. YOUR SOLUTION SL
This system is built by Machine learning and regression model. By using this model we can predict the resale value of the car at any time anywhere.

8. CHANNELS of BEHAVIOUR CH
8.1 ONLINE
When researching, customers don't look for information on auto brand websites alone, they visit comparison sites to check prices and user reviews.

4. EMOTIONS: BEFORE / AFTER EM
Before buying a car they experience a state of elation, joy, delight etc.,

After buying a car the sense of achievement they feel when they drive it home is linked to the fact that you are now the owner of the car.

8.2 OFFLINE
When customer wanted to buy a car they would visit one auto dealership after another, talking with salespeople and seeing where they could get the best price.

13

4. **REQUIREMENT ANALYSIS**

4.1 FUNCTIONAL REQUIREMENTS

The functional requirements for the suggested solution are listed below.

FR No.	Functional Requirement	Sub Requirement
FR-1	User Registration	Using a Form to Register using Gmail to sign up using LinkedIN to sign up
FR-2	User Confirmation	Email confirmation required Confirmation via OTP
FR-3	User Profile	User Details
FR-4	Database	Car Database Customer Database
FR-5	Features and technology	Performance of the car , fuel capacity , mileage etc.,
FR-6	Feedback	Feedback through Form Feedback through Gmail Feedback through LinkedIN

4.2 NON-FUNCTIONAL REQUIREMENTS

The non-functional requirements for the suggested solution are listed below.

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Excellent user interface, Precise value prediction
NFR-2	Security	Securing user passwords and personal information
NFR-3	Reliability	Failure occurs less frequently and is failure free.
NFR-4	Performance	execute accurate prediction value, The landing page must accommodate multiple users and have a response time of no more than five seconds.
NFR-5	Availability	Services must always be available without interruption, with the exception of when servers are being updated.
NFR-6	Scalability	It can instantaneously serve millions of users located all over the world and manage any amount of data and computations in a time- and money- saving manner.

5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAM
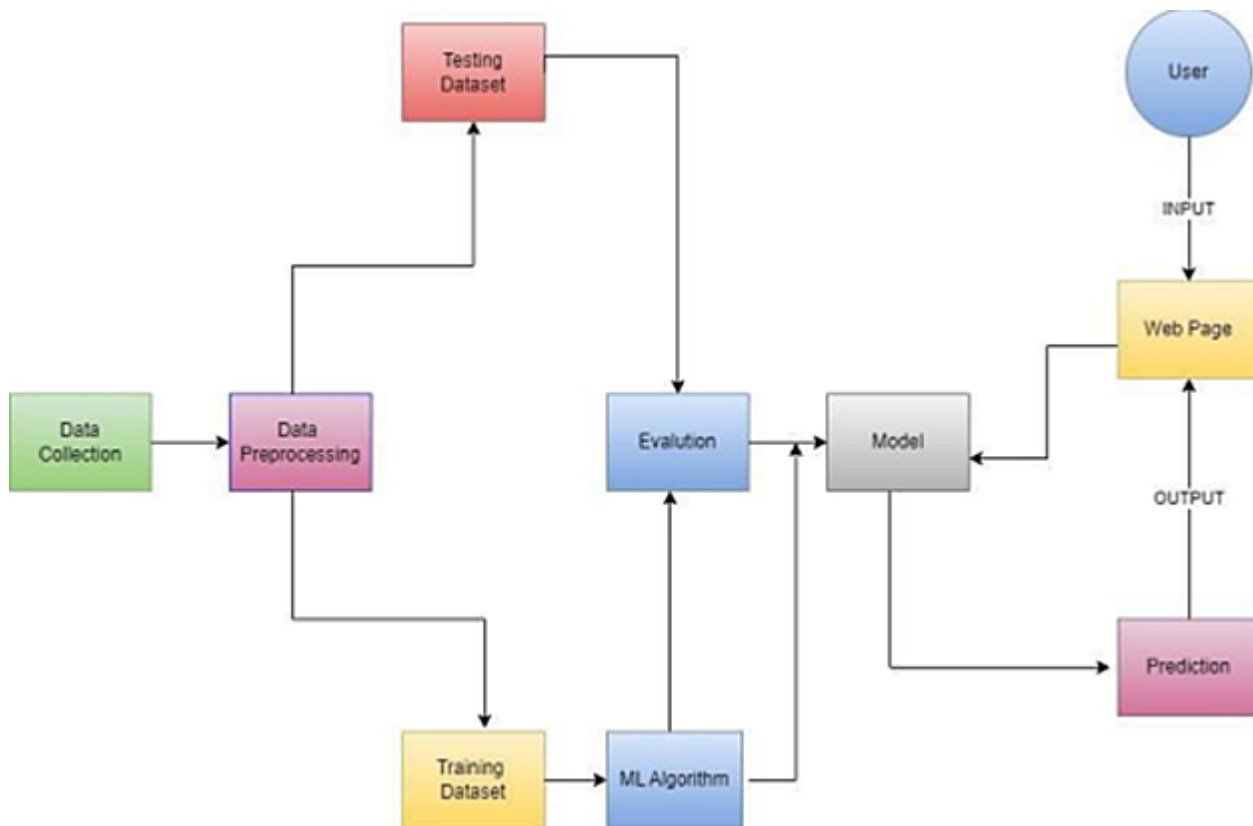
5.2 SOLUTION & TECHNICAL ARCHITECTURE
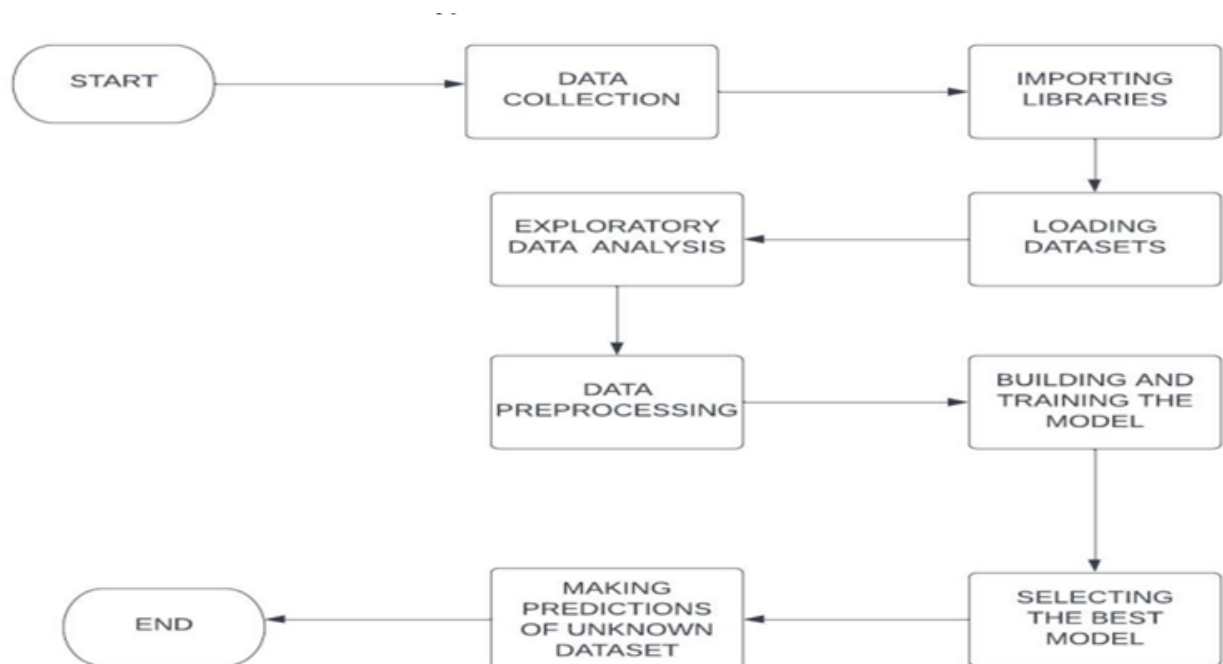
5.3 USER STORIES

User type	Functional requirement (epic)	User Story Number	User Story/Task	Acceptance criteria	Priority	release
Customer (mobile user)	Data entry	USN-1	I can add information about my car to the application as a user.	I can input the automobile information.	Medium	Sprint-1
Customer (mobile user)	Obtain output	USN-2	I will be given access to the application's car resale value as a user.	I'm able to get my car's resale value.	High	Sprint-1
Customer (mobile user)	Data entry	USN-1	I can enter the car's information into the application as a user.	I can enter the car details	Medium	Sprint-1
Customer (mobile user)	Obtain output	USN-2	I will be given access to the application's car resale value as a user.	I'm able to get my car's resale value.	High	Sprint-1

6. **PROJECT PLANNING & SCHEDULING**

6.1 SPRINT PLANNING & ESTIMATION

Below template shows product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can enter into the website with the help of the Google chrome browser in Windows	2	High	4
Sprint-1	Registration	USN-2	As a user, I can enter into the website through browser in Android	1	High	4

Sprint-1	Registration	USN-3	As a user, I can enter into the website through browser in ios	2	Medium	4
Sprint-1	Login	USN-4	As a user, I can find the car resale value prediction page on the website	1	High	4
Sprint-2	Home Page	USN-5	As a user, I need to select the parameters like Year, Showroom price, Kilometres driven, fuel type etc and click on the submit button	2	High	4

Sprint-3	Home Page	USN-6	As a user, I can see the accurate price for car resale after entering the details.	2	High	4
Sprint-4	Home Page	USN-7	As a user, If I made a mistake while providing the details, I can reset the details and click the submit button.	1	Low	4

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	20	6 Days	24 Oct 2022	29 Oct 2022	20	15 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	15 Oct 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	15 Oct 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	15 Oct 2022

6.3 REPORTS FROM JIRA

7. **CODING**
   7.1 FEATURE 1
         Predict the accurate results: It will predict the accurate results that will result
in the accurate value for resale.

   7.2 FEATURE 2
         Nice and smooth user interface and experience:'

8. TESTING

8.1 TEST CASES

a. Verify that get user data from register and pass it ot backend?

b. Verifiy the Output Display page can be reachable?

c. Verify all the UI elements in Output Display page rendered properly?

d. Verify user is able to get predicted result?
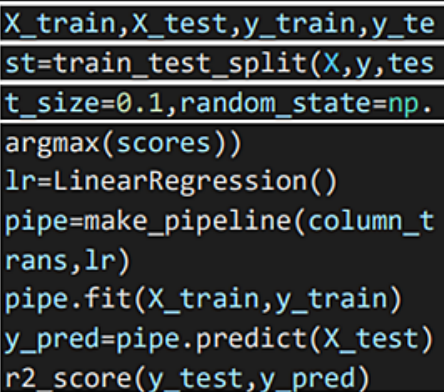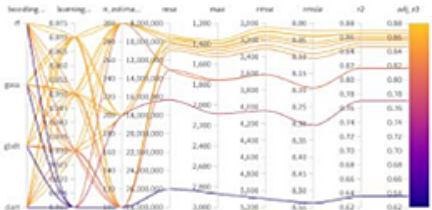
8.2 USER ACCEPTANCE TESTING

		Date	18-Nov-22	
		Team ID	PNT2022TMID31751	
		Project Name	Car Resale value Prediction	
		Maximum Marks	4 marks	

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID	Executed By
DataEntryPage_tst_cse1	Functional	Data Entry Page	Get user data from register and	Html,css,Csv files	1.Enter URL and click go	Maruti Suzuki	User should be able to enter all	Working as	Pass		Y		Kubera Prijesh D,
DataEntryPage_tst_cse2	Functional	Data Entry Page	Verifiy the Output Display page	Html,css,Csv files	1.Enter URL and click go	-	User should navigate to Output	Working as	Pass		Y		Kubera Prijesh D,
OutputDisplayPage_tst_cse1	Functional	Output Display Page	Verify all the UI elements in	Html,css,Csv	1.Enter URL and click go	cleaned_dataset.csv,pickle file	All the UI elements rendered	Working as	Pass		Y		Kubera Prijesh D,
OutputDisplayPage_tst_cse2	Functional	Output Display Page	Verify user is able to get predicted	Html,css,Csv	1.Enter URL and click go	cleaned_dataset.csv	Predited Car Resale Value is	Working as	Pass		Y		Kubera Prijesh D,

9. RESULT

9.1 PERFORMANCE METRICS

Model accuracy at training : 90%

Model accuracy at testing : 80%

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Regression Model:** R2 score - 0.9013820624340118	```X_train,X_test,y_train,y_te
st=train_test_split(X,y,tes
t_size=0.1,random_state=np.
argmax(scores))
lr=LinearRegression()
pipe=make_pipeline(column_t
rans,lr)
pipe.fit(X_train,y_train)
y_pred=pipe.predict(X_test)
r2_score(y_test,y_pred)``` 0.9013820624340118
2.	Tune the Model	**Hyperparameter Tuning -** **Validation Method -**	

10. ADVANTAGES & DISADVANTAGES

The large number of factors that must be taken into account for an effective prediction makes predicting car prices a difficult undertaking. This project makes extremely accurate pricing predictions.With very little information collection, all car models are available and taken into consideration. The biggest benefit is that every sale has a correct value projected.

The process of predicting the resale value of a car requires more time and space because it is a difficult undertaking, and real-time data is not employed.

11. CONCLUSION

Car price prediction can be a challenging task due to the high number of attributes that should be considered for the accurate prediction. The major step in the 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.

Data cleaning is one of the processes that increases prediction performance, yet insufficient for the cases of complex data sets as the one in this research. Applying single machine algorithm on the data set accuracy was less than 50%. Therefore, the ensemble of multiple machine learning algorithms has been proposed and this combination of ML methods gains accuracy of 92.38%. This is significant improvement compared to single machine learning method approach. However, the drawback of the proposed system is that it consumes much more computational resources than single machine learning algorithm.

Although, this system has achieved astonishing performance in car price prediction problem our aim for the future research is to test this system to work successfully with various data sets.

12. FUTURE SCOPE

In the future, more data will be collected using different web-scraping techniques, and deep learning classifiers will be tested. Algorithms like Quantile Regression, ANN and SVM will be tested. Afterwards, the intelligent model will be integrated with web and mobile-based applications for public use. Moreover, after the data collection phase Semiconductor shortages have incurred after the pandemic which led to an increase in car prices, and greatly affected the secondhand market. Hence having a regular Data collection and analysis is required periodically, ideally, we would be having a real time processing program

13. APPENDIX
13.1 SOURCE CODE

app.py:

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

app = Flask(__name__)

model = pickle.load(open(r'LinearRegressionModel.pkl', 'rb'))
car = pd.read_csv('Cleaned_dataset.csv')


@app.route('/')
def index():
    companies = sorted(car['Brands'].unique())
    car_models = sorted(car['Car_names'].unique())
    year = sorted(car['year'].unique(), reverse=True)
    fuel_type = car['fuel_type'].unique()

    companies.insert(0, 'Select Company')
    return render_template('index.html', companies=companies,
car_models=car_models, years=year, fuel_types=fuel_type)


@app.route('/predict', methods=['POST'])
def predict():
    company = request.form.get('company')

    car_model = request.form.get('car_models')
    year = request.form.get('year')
    fuel_type = request.form.get('fuel_type')
    driven = request.form.get('kilo_driven')

    prediction = model.predict(pd.DataFrame(columns=['name', 'company',
'year', 'kms_driven', 'fuel_type'],
                                            data=np.array([car_model,
company, year, driven, fuel_type]).reshape(1, 5)))
    print(prediction)
```

```
    return str(np.round(prediction[0], 2))



if __name__ == '__main__':
    app.run()
```

car_resale_value_prediction.ipynb:

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import *
from sklearn.model_selection import *
from sklearn.ensemble import *
from sklearn.feature_selection import *
from sklearn.metrics import *
from sklearn.tree import *
from sklearn.svm import *
from sklearn.preprocessing import LabelEncoder
import pickle
import warnings

warnings.filterwarnings('ignore')
%matplotlib inline
mpl.style.use('ggplot')
data = pd.read_csv('Datasets_of_cars.csv')

data.head()
data.shape
data.info()
backup=data.copy()
data['year'].unique()
data['Price'].unique()

```python
data['kms_driven'].unique()
data['fuel_type'].unique()
data['Car_names'].unique()
#year has many non-year values
data=data[data['year'].str.isnumeric()]
#year is in object. Change to integer
data['year']=data['year'].astype(int)
#Price has Ask for Price
data=data[data['Price']!='Ask For Price']
#Price has commas in its prices and is in object
data['Price']=data['Price'].str.replace(',','').astype(int)
#kms_driven has object values with kms at last
data['kms_driven']=data['kms_driven'].str.split().str.get(0).str.replace(',','')
#It has nan values and two rows have 'Petrol' in them
data=data[data['kms_driven'].str.isnumeric()]
data['kms_driven']=data['kms_driven'].astype(int)
#fuel_type has nan values
data=data[~data['fuel_type'].isna()]
data.shape
data['Car_names']=data['Car_names'].str.split().str.slice(start=0,stop=3).str.join(' ')
#Resetting the index of the final cleaned data
data=data.reset_index(drop=True)
data
data.to_csv('Cleaned_datasets.csv')
data.info()
data.describe()
data['Brands'].unique()
data.hist(figsize=(8,8))
sns.pairplot(data)
sns.boxplot(data.Price)
Extracting Training Data
X=data[['Car_names','Brands','year','kms_driven','fuel_type']]
y=data['Price']
X
y.shape
#Encoding Techniques
le=LabelEncoder()
data.head()
plt.figure(figsize=(10,8))
sns.heatmap(data.corr(),annot=True)
# dependent variable
```

```python
y=data['Price']
y
```
#independent variable
```python
X=data.drop(columns=['Price'],axis=1)
X.head()
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2)
X_train.shape
y_train.shape
X_test.shape
y_test.shape
X = data.drop(columns='Price')
y = data['Price']
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2)
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import make_column_transformer
from sklearn.pipeline import make_pipeline
from sklearn.metrics import r2_score
enc=OneHotEncoder()
enc.fit(X[['Car_names','Brands','fuel_type']])
column_trans=make_column_transformer((OneHotEncoder(categories=enc.categories
_),['Car_names','Brands','fuel_type']),
                        remainder='passthrough')
# Linear Regression model
mod=LinearRegression()
```
 # Making a pipeline
```python
pipe=make_pipeline(column_trans,mod)
```
 # Fitting the model
```python
pipe.fit(X_train,y_train)
```
#Predicting the values to test
```python
y_pred=pipe.predict(X_test)
y_pred
```
#printing the accuracy
```python
r2_score(y_test,y_pred)
scores=[]
for i in range(1000):
    X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.1,random_state=i)
    mod=LinearRegression()
    pipe=make_pipeline(column_trans,mod)
```

```
    pipe.fit(X_train,y_train)
    y_pred=pipe.predict(X_test)
    scores.append(r2_score(y_test,y_pred))
```

np.argmax(scores)

scores[np.argmax(scores)]

pipe.predict(pd.DataFrame(columns=X_test.columns,data=np.array(['Maruti Suzuki Swift','Maruti',2019,100,'Petrol']).reshape(1,5)))

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.1,random_state=np.argmax(scores))

lr=LinearRegression()

pipe=make_pipeline(column_trans,lr)

pipe.fit(X_train,y_train)

y_pred=pipe.predict(X_test)

#R2score
r2_score(y_test,y_pred)

 # Saving the Model to a Pickle File
pickle.dump(pipe,open('LinearRegressionModel.pkl','wb'))
pipe.predict(pd.DataFrame(columns=X_test.columns,data=np.array(['Maruti Suzuki Swift','Maruti',2019,100,'Petrol']).reshape(1,5)))
pipe.steps[0][1].transformers[0][1].categories[0]

13.2 GITHUB & PROJECT DEMO LINK
Github link : https://github.com/IBM-EPBL/IBM-Project-47160-1660796890

Demo Link: https://drive.google.com/file/d/14cUo6FRDwcC5yoFsDdF3qmjFbS9OciA/view?usp=share_link