CAR RESALE VALUE PREDICTION

TEAM ID : PNT2022TMID37881

SI.NO	TABLE OF CONTENTS	PAGE NO.
1.	**INTRODUCTION** 1.1 Project Overview 1.2 Purpose	4 4
2.	**LITERATURE SURVEY** 2.1 Existing Problem 2.2 References 2.3 Problem Statement Definition	18 18 20
3.	**IDEATION AND PROPOSED SOLUTION** 3.1 Empathy Map Canvas 3.2 Ideation & Brainstorming 3.3 Proposed Solution 3.4 Proposed Solution Fit	21 21 22 31
4.	**REQUIREMENT ANALYSIS** 4.1 Functional Requirements 4.2 Non-Functional Requirements	32 33
5.	**PROJECT DESIGN** 5.1 Data Flow Diagrams 5.2 Solution & Technical Architecture 5.3 User Stories	35 36 37
6.	**PROJECT PLANNING & SCHEDULING** 6.1 Sprint Planning & Estimation 6.2 Sprint Delivery Schedule 6.3 Reports from JIRA	38 38 39
7.	**CODING AND SOLUTIONING (Explain the features added in the project along with code)** 7.1 Feature 1 7.2 Feature 2	40 40

1. INTRODUCTION

1.1 Project Overview

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models across cities in the United States. Our results show that Random Forest model and K-Means clustering with linear regression yield the best results, but are compute heavy. Conventional linear regression also yielded satisfactory results, with the advantage of a significantly lower training time in comparison to the aforementioned methods

1.2 Purpose

Deciding whether a used car is worth the posted price when you see listings online can be difficult. Several factors, including mileage, make, model, year, etc. can influence the actual worth of a car. From the perspective of a seller, it is also a dilemma to price a used car appropriately[2-3]. Based on existing data, the aim is to use machine learning algorithms to develop models for predicting used car prices.

2. LITERATURE SURVEY

SI No.	TITLE	JOURNAL	AUTHOR	CHALLENGES/ FUTURE SCOPE
1.	Used car price prediction	IRJET	praful rana, deep pandiya, dhawal kotak	n future this machine learning model may bind with various website which can provide real time data for price prediction. Also we may add large historical data of car price which can help to improve accuracy of the machine learning model. We can build an android app as user interface for interacting with user. For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train

				on clusters of data rather than the whole dataset.
2.	used car price prediction and life span	IARJSET	aditya nikhade , rohan borde	This Project In machine learning model that will be connected with may dataset and with various website which can provide real time data for price prediction Will Stored in their site or GitHub. Also, we may add big amount of data of car price which can help an improve accuracy of the machine learning model . We also trying to develop an android app as user interface for interacting and user friendly with user. For better performance of the model, we also plan a to use neural network.

3.	vehicle resale price prediction using machine learning	Juni Khyat (UGC Care Group I Listed urnal)	B.Lavanya, Sk.Reshma, N.Nikitha, M.Namitha, L.Kanya Kumar, S.Kishore Babu,	In this paper, four distinctive AI procedures have been utilized to figure the cost of pre-owned vehicles in Mauritius. The mean blunder with direct relapse was about Rs 51,000 while for kNN it was about Rs 27,000 for Nissan vehicles and about Rs 45,000 for Toyota vehicles. J48 and Naïve Bayes exactness hung between 60-70% for various blends of boundaries. The primary shortcoming of choice trees and credulous bayes is their powerlessness to deal with yield classes with numeric qualities. Consequently, the value quality must be ordered into classes which contained a

				scope of costs yet this clearly presented further justification for errors. The primary limit of this examination is the low number of records that have been utilized. As future work, we plan to gather more information and to utilizes further developed methods like counterfeit neural organizations, fluffy logic and hereditary calculations to foresee vehicle costs.
4.	Predicting Used Car	CS 229 Project Report	Kshitij Kumbar,	For better performance, we plan

	Prices		Pranav Gadre and Varun Nayak	to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset. To correct for overfitting in Random Forest, different selections of features and number of trees will be tested to check for change in performance.

5.	Used Cars Price Prediction using Supervised Learning Techniques	International Journal of Engineering and Advanced Technology	Mukkesh Ganesh	The prediction error rate of all the models was well under the accepted 5% of error. But, on further analysis, the mean error of the regression tree model was found to be more than the mean error rate of the multiple regression and lasso regression models. Even though for some seeds the regression tree has better accuracy, its error rates are higher for the rest. This has been confirmed by performing an ANOVA. Also, the post-hoc test revealed that the error rates in multiple regression models and lasso regression models aren't significantly different from each

				other. To get even more accurate models, we can also choose more advanced machine learning algorithms such as random forests, an ensemble learning algorithm which creates multiple decision/regression trees, which brings down overfitting massively or Boosting, which tries to bias the overall model by weighing in the favor of good performers. More data from newer websites and different countries can also be scraped and this data can be used to retrain these models to check for reproducibility.

6.	predictive analysis of used car prices using machine learning	International Research Journal of Modernization in Engineering Technology and Science	Ashutosh Datt Sharma ,Vibhor Sharma,Sahil Mittal,Gautam Jain,Sudha Narang	Predicting prices of a used car is a challenging task because of a high number of features and parameters that should be considered to generate accurate results. The first and foremost step is data gathering and preprocessing data. Then a model was defined and created for implementing algorithms and generating results. After applying various regression algorithms on the model, it could be concluded that Decision Tree Algorithm was the best performer with highest r2 score of 0.95 which simply signified the fact that it generated the most accurate

				predictions as reflected by the Original v/s Prediction line graph. Apart from a best r2 score, Decision Tree also had the least Mean Squared Error and Root Mean Squared Values that shows that the errors in predictions were least among all and therefore the results generated are highly accurate. .

7.	Price Prediction for Used Cars	Mid Sweden University.	Marcus Collard	the best potential for development of a consumer tool for evaluating used cars or a particular subset of used cars. The results show that Random Forest Regression performed the best on all performance metrics and for all price percentile subsets of used cars. It was also much better able to approximate the depreciation.
8.	Car Price Prediction using Machine Learning Techniques	TEM Journal. Volume 8	Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric	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.
9.	Used Cars Price Prediction and Valuation using Data Mining Techniques	Rochester Institute of Technology	Abdulla AlShared	Using data mining and machine learning approaches, this project proposed a scalable framework for Dubai based used cars price prediction. Buyanycar.com website was scraped using the Parse Hub scraping tool to collect the benchmark data. An efficient machine learning model is built by training, testing, and evaluating three machine learning regressors named Random Forest Regressor, Linear Regression, and

				Bagging Regressor. As a result of preprocessing and transformation, Random Forest Regress or came out on top with 95% accuracy followed by Bagging Regress or with 88%. Each experiment was performed in realtime within the Google environment. In comparison to the system's integrated Jupiter notebook and Anaconda's platform, algorithms took less training time in Google .
10.	Consumer preferences for electric vehicles: a Consumer preferences for electric vehicles: a	Transport Reviews	Fanchao Liao, Eric Molin , Bert van Wee	In general, the effect of individualspecific variables on EV preference remains an open question. Psychological variables are the exception and have a

				proven stable effect, shown by several studies. For socioeconomic and demographic variables, the impact is unclear and sensitive to small changes in model specification. The direction of the effect is also ambiguous since existing evidence is contradictory. Other variables are only included in a few studies, therefore their effects are as yet inconclusive. In most cases, the correlation between all these variables has not been controlled for to avoid self - selection bias. More research is definitely necessary to clarify these currently fuzzy relationships and other methods are

				needed to add more and confidence to the results

2.1 Existing Problem

The real reason that this problem exist is in this car resale value prediction system cant predict exact price as brand owners price. This just predicts approx. the value by interior and exterior, bs4 and bs6, petrol or diesel.

2.2 References

[1] NATIONAL TRANSPORT AUTHORITY. 2014. Available from: http://nta.gov.mu/English/Statistics/Pages/Archive.aspx [Accessed 15 January 2014].

[2] MOTORS MEGA. 2014. Available from: http://motors.mega.mu/news/2013/12/17/auto-market-8-decrease-sales-newcars/ [Accessed 17 January 2014].

[3] LISTIANI, M., 2009. Support Vector Regression Analysis for Price Prediction in a Car Leasing Application. Thesis (MSc). Hamburg University of Technology.

[4] RICHARDSON, M., 2009. Determinants of Used Car Resale Value. Thesis (BSc). The Colorado College.

[5] WU, J. D., HSU, C. C. AND CHEN, H. C., 2009. An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications. Vol. 36, Issue 4, pp. 7809-7817.

[6] DU, J., XIE, L. AND SCHROEDER S., 2009. Practice Prize Paper - PIN Optimal Distribution of Auction Vehicles System: Applying Price Forecasting, Elasticity Estimation and Genetic Algorithms to Used-Vehicle Distribution. Marketing Science, Vol. 28, Issue 4, pp. 637-644.

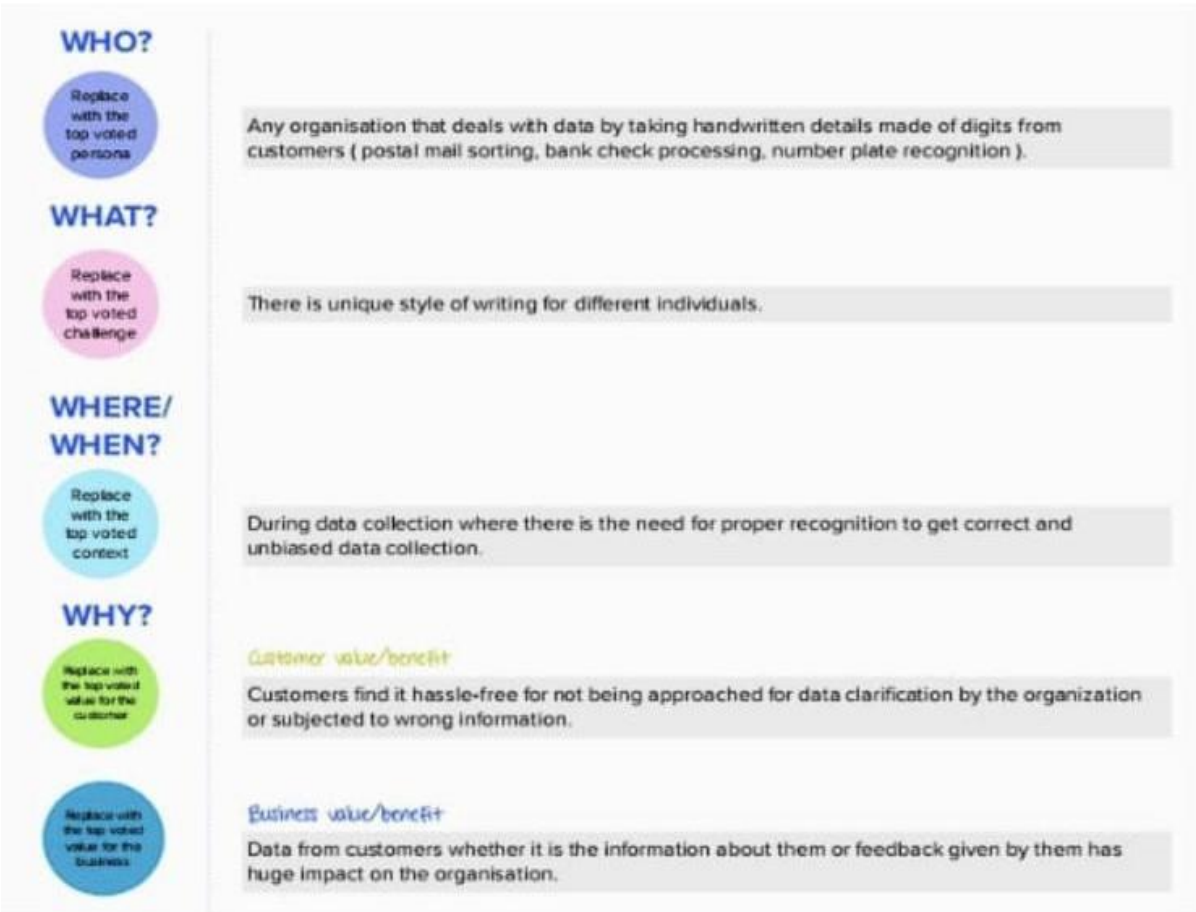
[7] GONGGI, S., 2011. New model for residual value prediction of used cars based on BP neural network and non-linear curve fit. In: Proceedings of the 3 rd IEEE International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Vol 2. pp. 682-685, IEEE Computer Society, Washington DC, USA.

[8] LEXPRESS.MU ONLINE. 2014. Available from: http://www.lexpress.mu/ [Accessed 17 January 2014]

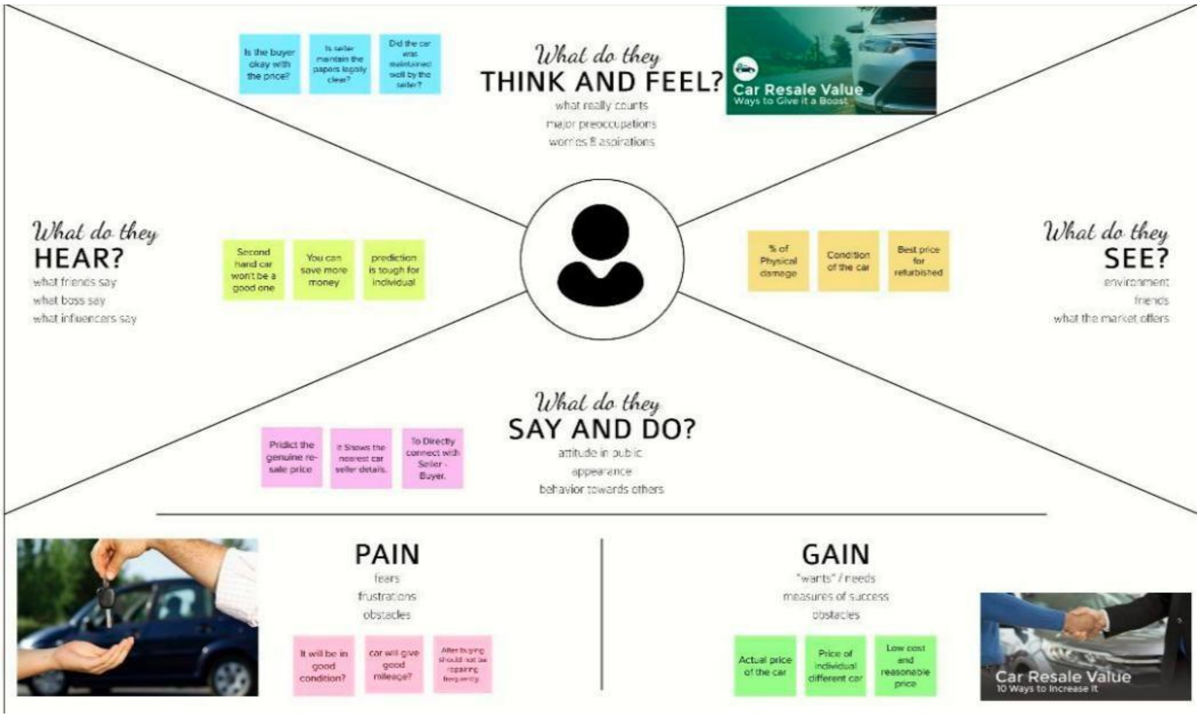
[9] LE DEFI MEDIA GROUP. 2014. Available from: http://www.defimedia.info/ [Accessed 17 January 2014]

[10] GELMAN, A. AND HILL, J., 2006. Data Analysis Using Regression and Multilevel Hierarchical Models. Cambridge University Press, New York, USA.

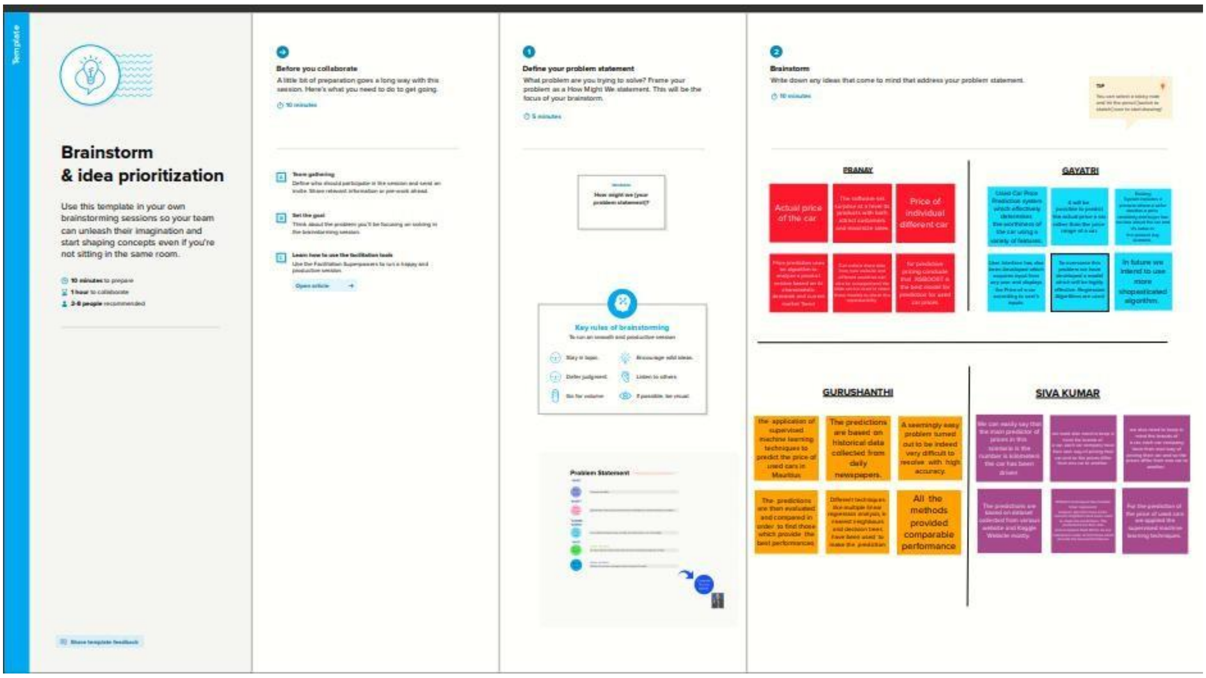
2.3 Problem Statement Definition



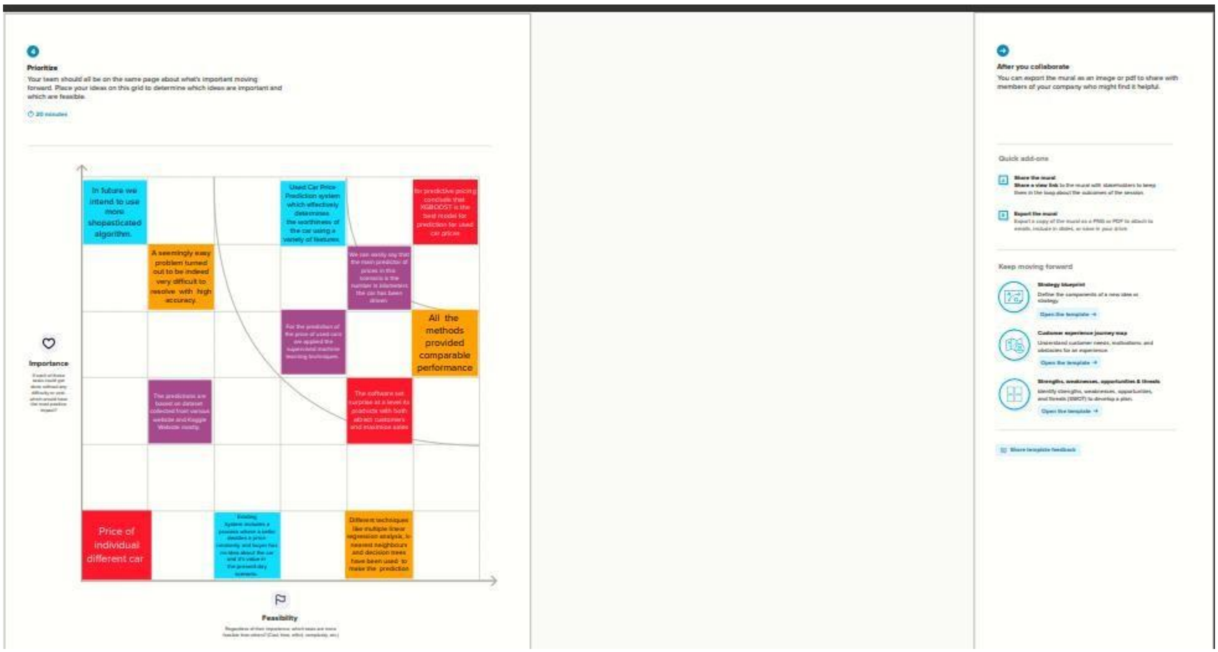
3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming



3.3 Proposed Solution



S.No.	Parameter	Decription
1.	Problem Statement (Problem to be solved)	The real reason that this problem exist is in this car resale value prediction system cant predict exact price as brand owners price. This just predicts approx. the value by interior and exterior, bs4 and bs6, petrol or diesel.
2.	Idea / Solution description	Building a software with sensor that can scan the car totally and calculate the rate of damage and condition of car

		accurately and predict the price as per the car condition and damage. Comparison and price prediction of different brands.
3.	Novelty / Uniqueness	To give customers a Standard and Friendly service which would make them feel comfortable
4.	Social Impact / Customer Satisfaction	The role of cars has become highly important, though controversial. They are used throughout the world and have become the most popular mode of transport in many of the more developed countries. In developing countries, the effects of the car on society are not as visible, however they are nonetheless significant. The development of the car built upon the transport

		sector first started by railways. This has introduced sweeping changes in employment patterns, social interactions, infrastructure and the distribution of goods. Despite the positive effects on access to remote places and mobility, comfort provided by the automobile, allowing people to geographically increase their social and economic interactions, the negative effects of the car on everyday life are not negligible. Although the introduction of the massproduced car represented a revolution in industry and convenience. Creating job demand and tax revenue, the high

		motorisation rates also brought severe consequences to the society and to the environment. The modern negative consequences of heavy automotive use include the use of non-renewable fuels, a dramatic increase in the rate of accidental death, the disconnection of local community, the decrease of local economy, the rise in obesity and cardiovascular diseases, the emission of air and noise pollution, The emission of greenhouse gases, generation of urban sprawl and traffic, segregation of pedestrians and other active mobility means of transport, decrease in the railway network, urban decay and the high cost

		per unit-distance on which the car paradigm is based.
5.	Business Model (Revenue Model)	Types of Revenue Streams There are several ways to generate Revenue Streams: Asset sale The most widely understood Revenue Stream derives from selling ownership rights to a physical product. Amazon.com sells books, music, consumer electronics, and more online. Fiat sells automobiles, which buyers are free to drive, resell, or even destroy. Usage fee This Revenue Stream is generated by the use of a particular service. The more a service is used, the more the customer pays. A telecom operator may charge customers for the number of minutes spent

		on the phone. A hotel charges customers for the number of nights rooms are used. A package delivery service charges customers for the delivery of a parcel from one location to another. Subscription fees This Revenue Stream is generated by selling continuous access to a service. A gym sells its members monthly or yearly subscriptions in exchange for access to its exercise facilities. World of Warcraft Online, a Web-based computer game, allows users to play its online game in exchange for a monthly subscription fee. Nokia's Comes with Music service gives users access to a music library for a subscription fee. Lending/Renting/Leasing

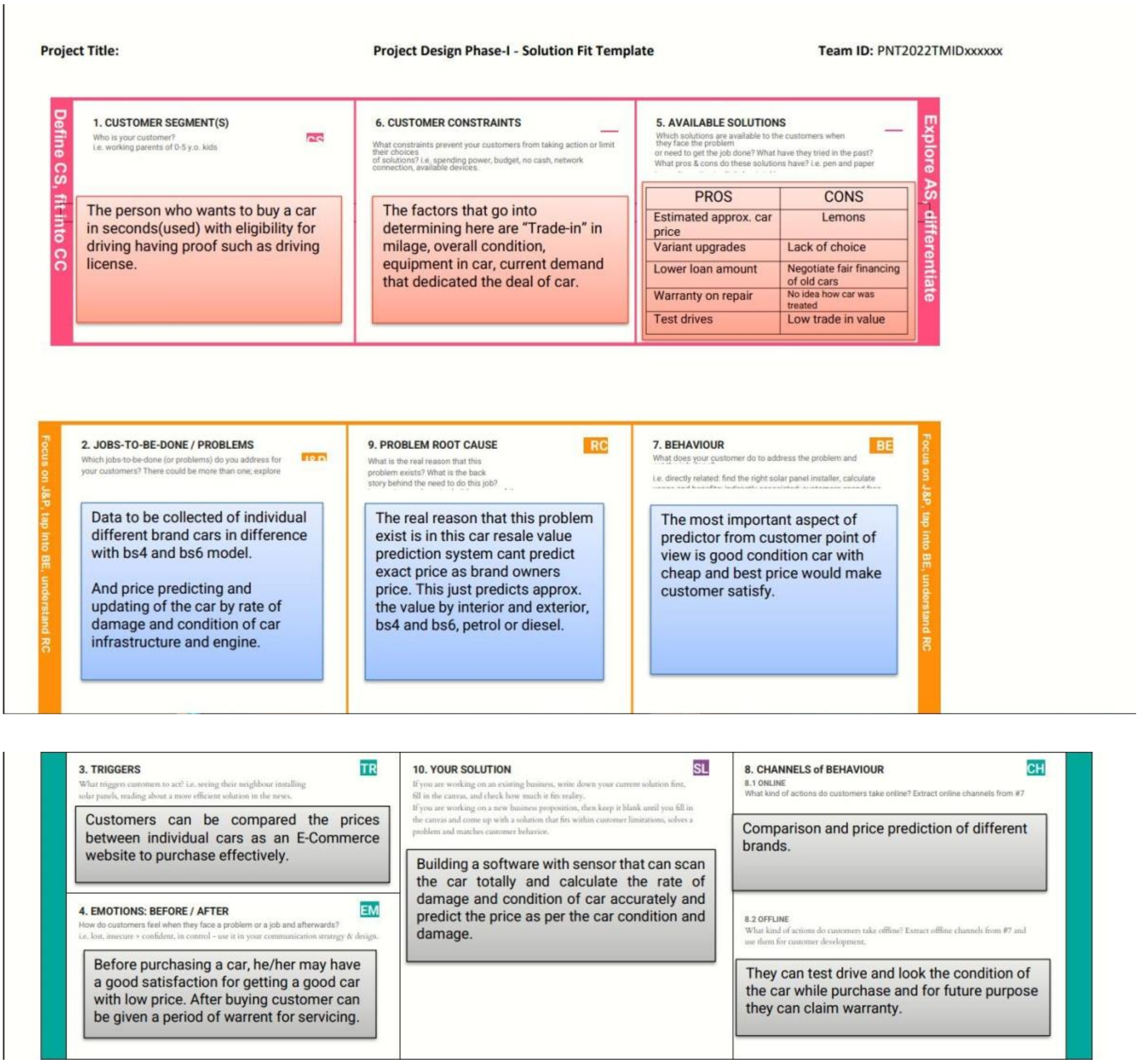
		This Revenue Stream is created by temporarily granting someone the exclusive right to use a particular asset for a fixed period in return for a fee. For the lender this provides the advantage of recurring revenues. Renters or lessees, on the other hand, enjoy the benefits of incurring expenses for only a limited time rather than bearing the full costs of ownership. Zipcar.com provides a good illustration. The company allows customers to rent cars by the hour in North American cities. Service has led many people to decide to rent rather than purchase automobiles. Licensing This Revenue Stream is generated by giving customers

		permission to use protected intellectual property in exchange for licensing fees. Licensing allows rights holders to generate revenues from their property without having to manufacture a product or commercialize a service. Licensing is common in the media industry, where content owners retain copyright while selling usage licenses to third parties. Similarly, in technology sectors patent holders grant other companies the right to use a patented technology in return for a license fee. Brokerage fees This Revenue Stream derives from intermediation services performed on behalf of two or more parties. Credit card providers, for

		example, earn revenues by taking a percentage of the value of each sales transaction executed between credit card merchants and customers. Brokers and real estate agents earn a commission each time they successfully match a buyer and seller. Advertising This Revenue Stream results from fees for advertising a particular product, service, or brand. Traditionally, the media industry and event organizers relied heavily on revenues from advertising. In recent years other sectors, including software and services, have started relying more heavily on advertising revenues.
6.	Scalability of the Solution	Pre-owned vehicle ecommerce business

		replicates MySQL data, saves six engineers over four months of manual work & improves data reliability for analytics teams

3.4 Proposed Solution Fit



4. REQUIREMENT ANALYSIS

4.1 Functional Requirements

FR No.	**Functional Requirement (Epic)**	**Sub Requirement (Story / Sub-Task)**
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	Core functionality	Recognize the human handwritten digits from different sources like images, papers, touch screens, etc, and classify them into 10 predefined classes (0-9)
FR-3	Access	Able to copy the recognised digits, Focus a part of the image manually.
FR-4	Network	The database has to be updated for training for more accuracy.

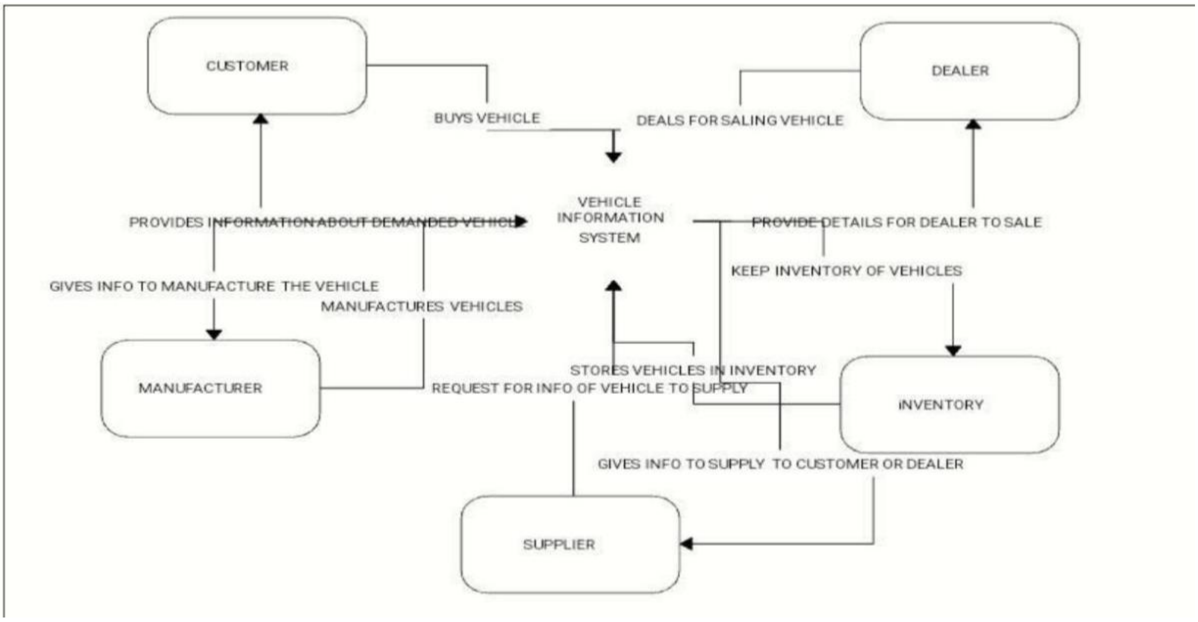
4.2 Non-Functional Requirements

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Recognising handwritten information such as reading postal addresses, bank check amounts, and forms.
NFR-2	Security	When the image is passed to recognise a particular area of digit(s), the image will not be stored at the backend.
NFR-3	Reliability	CNN has shown remarkable abilities in offline handwritten character recognition of Arabic language; handwritten Tamil character recognition; Telugu character recognition, handwritten Urdu text recognition, handwritten character recognition in Indic

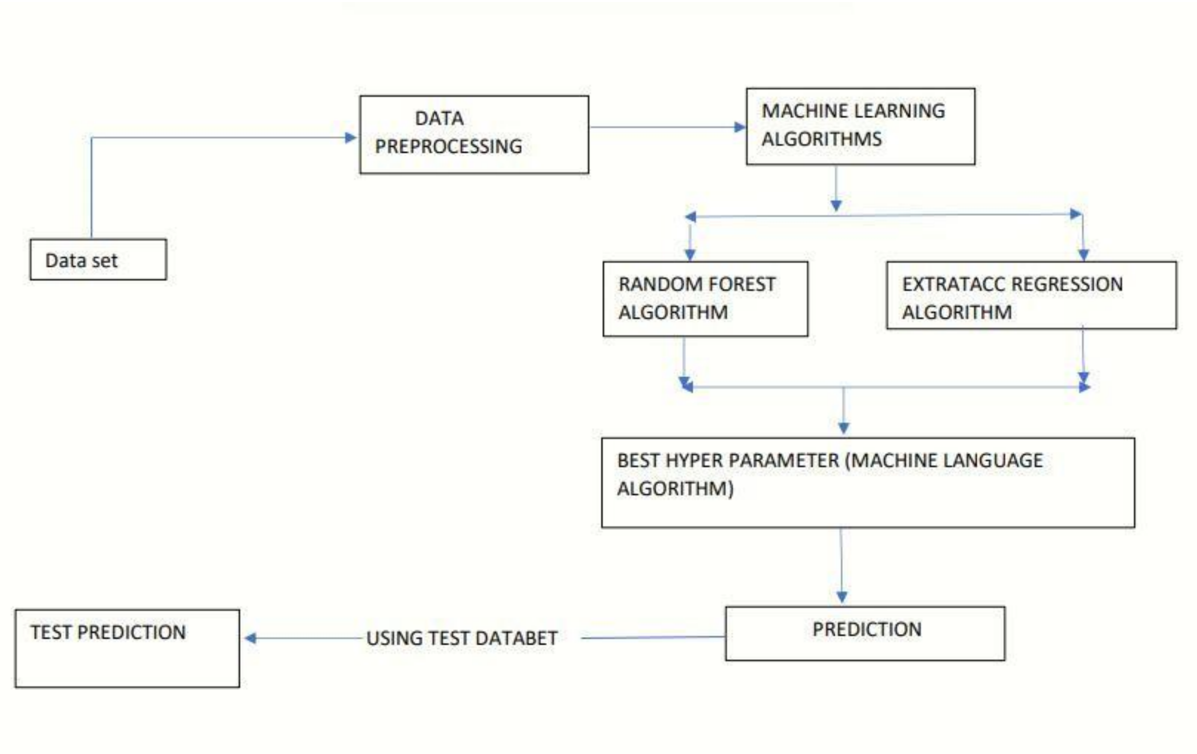
		scripts [44] and Chinese handwritten text recognition.
NFR-4	Performance	Hyper-parameters are, namely, activation function, number of epochs, kernel size, learning rate, hidden units, hidden layers, etc. that are responsible for the performance of the system.
NFR-5	Availability	There is no maintenance time separately for the servers to be down or can be accessed offline also.
NFR-6	Scalability	System will be such that it is easy to change, update, or add features later on.

5. PROJECT DESIGN

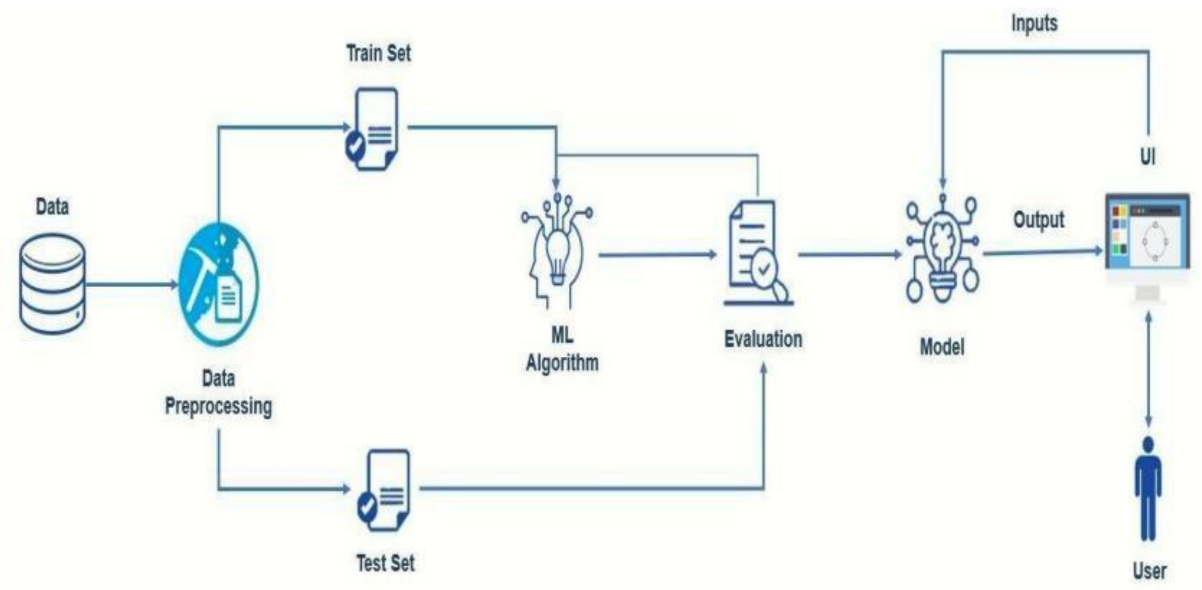
5.1 Data Flow Diagram



5.2 Solution Architecture



Technical Architecture:



5.3 User Stories:

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
			registration fee, insurance cost, repair work and general upkeep.			
	Create a Target List	USN-2	Once you have agreed on a budget, start making a list of requirements for your vehicle. You must also choose the type of vehicle you want. You can choose from SUVs, sedans, small cars and electric vehicles. It is recommended to check the reviews and ratings of the car you plan to purchase.		High	Sprint-1
	Research Your Options	USN-3	Used car dealerships are now presenting almost every corner of the city, everywhere in India. You can find the best dealer in town either by word of mouth or by comparing dealers online. Finding good dealers online is a fairly simple process. Just shortlist some popular second-hand car dealers and compare options available, cost, service and customer reviews before choosing the one for yourself.		Medium	Sprint-2
	Check the Vehicle's History	USN-4	Once you have explored various options and have narrowed down your search list, it is time to check the vehicle's health report. Check what kind of maintenance or repair works has it undergone. Double-check if the vehicle has ever been involved in a collision. If you are buying a used car in India, it is advisable to avoid buying a car that has been involved in an accident.		High	Sprint-1
	Call the Seller	USN-5	Contact the seller to double-check the information you have gathered about the vehicle. If you are buying from an individual seller, find out why they are selling the car and if there are any mechanical concerns. If you are considering a dealer, call to check the availability of the car. If everything goes fine, book an appointment for a test drive.		High	Sprint-1

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
	Test Drive	USN-6	A test drive will give you a clear idea about your shortlisted used car's condition. Take the car for a drive on different types of roads and cover a distance of at least four to five kilometers. You must also check the condition of the brakes and clutch while driving. Ensure that the speedometer and the distance recorder are working properly. If there is a vibration in the steering, it could mean some major issues with the engine.		High	
	Get a Professional to Inspect the Car	USN-7	When buying a used car, get a professional mechanic to inspect the car before you pay for it. If you buy a used car from a reputable dealer, the chances of receiving a damaged model are slim. Buying from a private seller, on the other hand, may necessitate a complete inspection by a skilled mechanic.		Medium	
	Double Check the Vehicle's Papers	USN-8	Before finalising the used car, it is advisable to check the papers properly. Check for the car's registration certificate; match the vehicle's engine number and chassis number. Check the insurance paper, PUC certificate along with the original sales invoice. This way, you can make sure the car you are buying is not stolen from its previous owner.		High	
	Negotiate Well	USN-9	This is when the real fun begins. Since you would have already set a budget for the car purchase, stick to it and negotiate with the seller over anything you deem important such as a major dent or bad paintwork. Since the cost of a used car is the seller's decision, make sure to negotiate well.		High	
	Used Car Finance	USN-10	Today, many financial institutions offer a loan for the purchase of used cars. If you are under a budget constraint, you may avail of this option. Before applying for a loan, compare the used car finance rates with different		Medium	

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
			lenders and check your used car loan eligibility with the lender of your choice. If you have a good profile and strong creditworthiness, you may seal a better deal on used car finance			
	Ownership Transfer	USN-11	The ownership of a car is transferred with its sale. The previous owner of the car must inform about the transfer to the RTO under which the vehicle is originally registered. This process must be initiated within 14 days along with a letter of intent and the details of the new owner.	I can access my account / dashboard	High	
Straight away	Drive Away	USN-12	Once you are done with the above formalities, it is time to announce your purchase and be a proud car owner. You can now spin off the car to your home or wherever the road calls you		High	

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Pre-process data	USN-1	Collect Dataset	1	Low	Pranay D
Sprint-1		USN-2	Import required libraries	1	Low	Gurushanthi
Sprint-1		USN-3	Read and clean data sets	2	Low	SivaKumar
Sprint-2	Model building	USN-1	Split data into independent and dependent variables	3	Medium	Gayatri
Sprint-2		USN-2	Apply using regression model	3	Medium	Pranay.D, Sivakumar
Sprint-3	Application building	USN-1	Build python flask application and HTML page	5	High	SivaKumar, Gurushanthi
Sprint-3		USN-2	Execute and test	5	High	Gayatri, Gurushanthi
Sprint-4	Training the model	USN-1	Train machine learning model	5	High	Pranay Gurushanthi
Sprint-4		USN-2	Integrate flask	5	High	Gayatri, Sivakumar

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

VELOCITY:

Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per unit).let’s calculate the team’s averge velocity (AV) per iteration unit (story points per day).

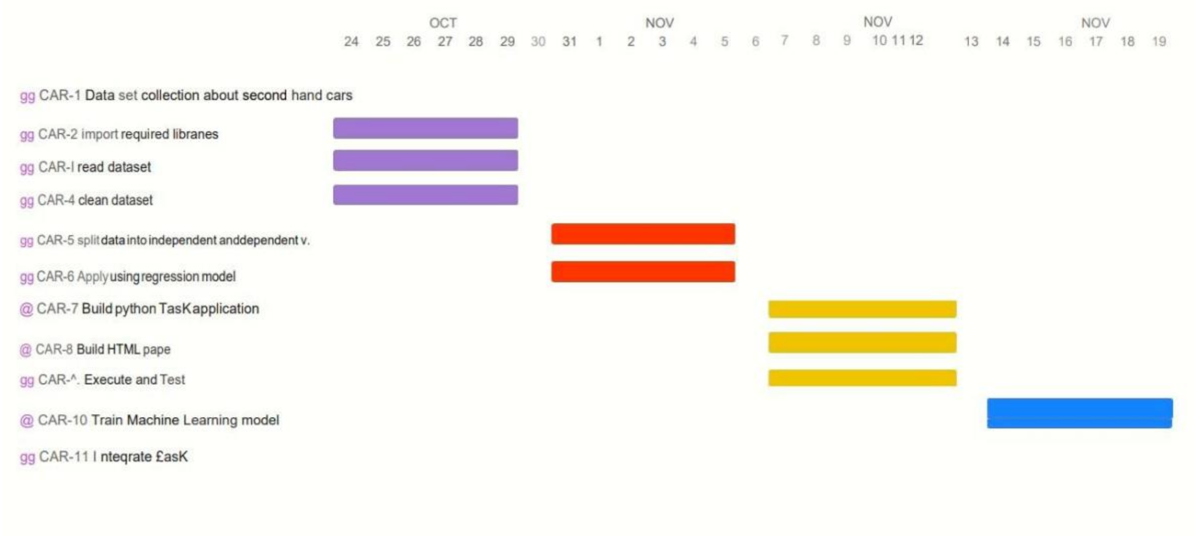
$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

6.3 Reports from JIRA

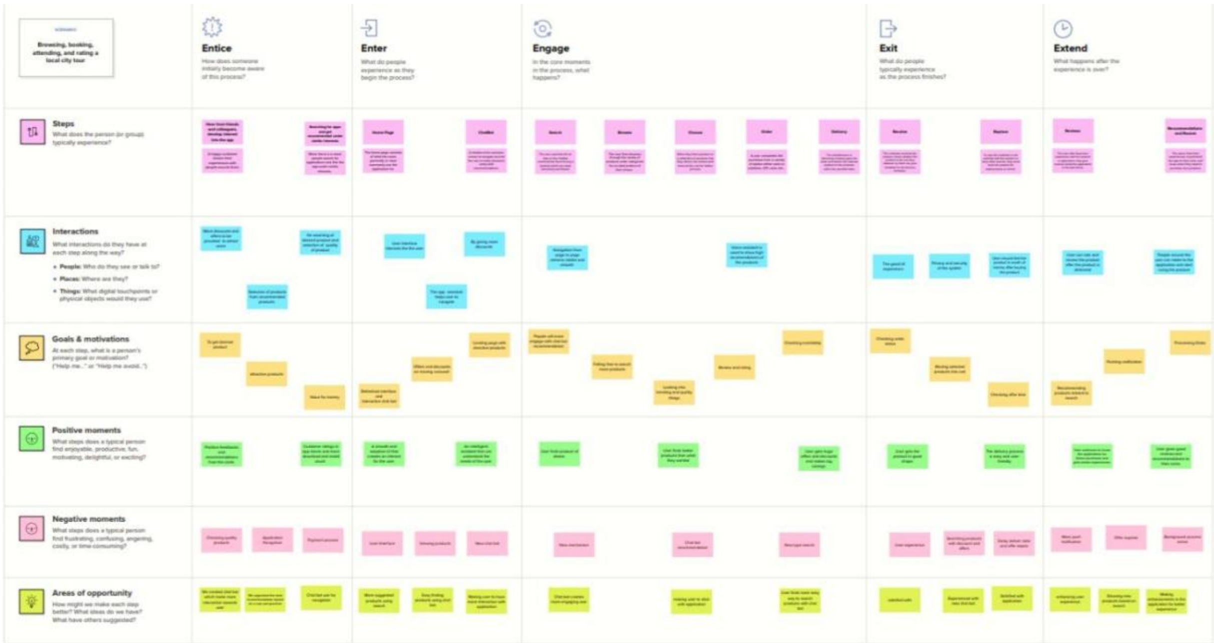
Burndown Chart:

A burndown chart is a graphical representation of work left to do versus time.It is often used in agile software development methodologies such as scrum. However, burn down charts can be applied to any project containing measurable progress over time.

SFRA	Sprint
SFRA-1 Registration	Sprint-1
SFRA-2 Verification	Sprint-1
SFRA-3 Login	Sprint-1
SFRA-4 Registration Alternative	Sprint-1
SFRA-5 Homepage	Sprint-2
SFRA-6 Result page	Sprint-2
SFRA-7 Create database for users	Sprint-2
SFRA-8 Feedback	Sprint-3
SFRA-9 Customer support- Chatbot	Sprint-4
SFRA-10 Testing and debugging the application	Sprint-5
SFRA-11 Containers of application	Sprint-5
SFRA-12 Deploy the application	Sprint-5



Customer Journey Map:



7. CODING AND 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

Date	03-Nov-22
Team ID	PNT2022TMID37881
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	Commnets	TC for Automation(Y/N)	BUG ID	Executed By
LoginPage_TC_OO1	login /sign in	Home Page	Verify user is able to see the Login/Signup popup when user clicked on My account button to see the webpage.	Network connection/ Available device for using website	1.Enter URL and click go 2.Click on My Account dropdown button 3.Verify login/Singup popup displayed or not	https://carprice.com/	Login/Signup popup should display	Working as expected	Pass				
LoginPage_TC_OO2	UI	Login page	Verify the UI elements in Login/Signup popup	Network connection / Available device for using website	1.Enter URL and click go 2.Click on My Account dropdown button 3.Verify login/Singup popup with below UI elements: a.email text box b.password text box c.Login button d.New customer? Create account link e.Last password? Recovery	https://resalevalueprice.com/	Application should show below UI elements: a.email text box b.password text box c.Login button d.New customer? Create account link e.Last password? Recovery password link	Working as expected	pass	Steps are clear to follow		BUG-1234	
LoginPage_TC_OO3	verification	Home page	Verify user is able to log into application with Valid credentials	Network connection ,Available device for using website, valid user name ,valid new password.	1.Enter URL(https://carprice.com/) and click go 2.Click on My Account dropdown button 3.purchase year 4. mainatance required	Username: preethi16266@gmail.com password: Testing123	User should navigate to user account homepage	successfully login	pass				
LoginPage_TC_OO4	availability	car model & brand	Available of car models & versions	Network connection ,Available device for using website, valid user name ,valid new password.	1.Enter URL(https://carprice.com/) and click go 2.Click on My Account dropdown button 3.choose the car model ad version 4.check the condition 5.accept the condition	Username: preethi1626@gmail password: Testing123	Application should show car model and resale predection value	shown	pass			BUG ID 234	
LoginPage_TC_OO4	resale value	Resale car value	Available resale car value and city of purchase	Network connection ,Available device for using website.	1.Enter URL (https://carprice.com/) and click go 2.Enter needed car model 3.Available model 4.Acutal price 5.Choose the car needed	Username: preethi1626@gmail.com password: Testing123678686786876876	Application should show model and resale predection value	shown	pass				
LoginPage_TC_OO5	type	Fule type	Verify the fuel content and Petrol or Disele and Mileage	Network connection ,Available device for using website	1.Enter URL (https://resalevalueprice.com/) and click go 2.Enter the fuel capasity 3.Enter the fuel type 4.choose the model of car and mileage	Username: preethi1626 password: Testing123678686786876876	Application should show the fule type and car model and mileage	shown	pass				
LoginPage_TC_OO6	machine verification	Transmission	verify the machine are automatic or non automatic	Network connection ,Available device for using website	1.Enter URL (https://resalevalueprice.com/) and click go 2.Enter the features 3.Enter the model type 4.choose the model	Username: preethi1626 password: Testing1236786867868768 7	Apllication shoul show the type of version automatic and non automatic	shown	pass				
LoginPage_TC_OO7	engine condition	Engine	verify the machine quality and condition	Network connection ,Available device for using website	1.Enter URL (https://resalevalueprice.com/) and click go 2.Enter the features of machine 3.Enter the machine model type 4.choose the machine condition	Username: preethi1626	Application shoul show the type of machine	shown	pass				
LoginPage_TC_OO8	resale values	car price	Choose the resale car price	Network connection ,Available device for using website	1.Enter URL (https://resalevalueprice.com/) and click go 2.Enter the features of car value price 3.Enter the resale price of car 4.choose the available car		Application should show the resale car price	shown	pass				

8.2 User Accept

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

Section	Total Cases	Not Tested	Fail	Pass
Login /sign in	1	0	0	1
User interface	1	0	0	1
Availability	1	0	0	1
type	1	0	0	1
condition	1	0	0	1
verification	2	0	0	2
Resale price	2	0	0	2

9. RESULTS

9.1 Performance Metrics

Car Resales Price Prediction

MODEL BUILDING

Choose the metrics of the model

```
#predicting the values to test set
y_pred = regressor.predict(X_test)
#printing the accuracy for test set
print(r2_score(Y_test,y_pred))
```

10. ADVANTAGES AND DISADVANTAGES

Price

Varient upgrades

Lower loan amount

Lower insurance premium

Warrenty premium

Buying from individuals

Higher intrest rates

Lack of choice

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. We will extend our test data with eBay [16] and OLX [17] used cars data sets and validate the proposed approach.

12. FUTURE SCOPE

For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset. To correct for overfitting in Random Forest, different selections of features and number of trees will be tested to check for change in performance.

13. APPENDIX

Source Code

```
# -- coding: utf-8 --
"""
Created on Thu Nov 17 13:44:45 2022


@author: pranay bharadwaj
"""



import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns


# TRAIN DATA PREPROCESSING
```

```
dataframe_train = pd.read_csv("E:/train-data.csv")

train_data = dataframe_train.copy()

for i in range(0, len(train_data)):
    if train_data['Power'][i] == 'null bhp':
        train_data['Power'][i] = np.nan

for i in range(0, len(train_data)):
    if train_data['Mileage'][i] == '0.0 kmpl' or train_data['Mileage'][i] == '0.0 km/kg':
        train_data['Mileage'][i] = np.nan

for i in range(0, len(train_data)):
    if train_data['Engine'][i] == 'null CC' or train_data['Engine'][i] == '0 CC':
        train_data['Engine'][i] = np.nan

train_data.drop(['New_Price'], axis=1, inplace=True)
train_data.drop(['Unnamed: 0'], axis=1, inplace=True)
```

```
train_data.dropna(inplace = True)

train_data.reset_index(inplace = True)

train_data.drop(['index'], axis=1, inplace=True)

y = train_data.iloc[:,-1].values

City = train_data['Location'].unique()

brand=[]

for i in range(0, 5844):
    k = train_data['Name'][i].split()
    brand.append(k[0].upper())

Brand = np.array(brand)

fig = plt.figure(figsize=(10,7))
fig.add_subplot(1,1,1)
```

```
ax = sns.countplot(Brand)
ax.set_xlabel("Brands")
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha='right')

Brand = pd.get_dummies(Brand, drop_first = True, dtype=int)

unique_brands=[]
for i in range(0,5844):
    if brand[i] in unique_brands:
        continue
    else:
        unique_brands.append(brand[i])

Loc = train_data['Location']

fig = plt.figure(figsize=(10,7))
fig.add_subplot(1,1,1)
ax = sns.countplot(Loc)
ax.set_xlabel("Location")
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha='right')
```

```
Loc = pd.get_dummies(Loc, drop_first = True, dtype=int)

train_data['Seats'] = train_data['Seats'].astype(int)

fig = plt.figure(figsize=(7,7))
fig.add_subplot(1,1,1)
ax = sns.countplot(train_data['Seats'])
ax.set_xlabel("Seats")

fig = plt.figure(figsize=(7,7))
fig.add_subplot(1,1,1)
ax = sns.countplot(train_data['Fuel_Type'])
ax.set_xlabel("Fuel Type")

fig = plt.figure(figsize=(7,7))
fig.add_subplot(1,1,1)
ax = sns.countplot(train_data['Transmission'])
ax.set_xlabel("Transmission")
```

```
fig = plt.figure(figsize=(7,7))
fig.add_subplot(1,1,1)
ax = sns.countplot(train_data['Owner_Type'])
ax.set_xlabel("Owner Type")

train_data.replace({'First': 1, 'Second': 2, 'Third': 3, 'Fourth & Above':
4}, inplace = True)

for i in range(0, 5844):
    k = train_data['Mileage'][i].split()
    train_data['Mileage'][i] = k[0]

for i in range(0, 5844):
    k = train_data['Power'][i].split()
    train_data['Power'][i] = k[0]

for i in range(0, 5844):
    k = train_data['Engine'][i].split()
    train_data['Engine'][i] = k[0]
```

```
train_data['Engine'] = train_data['Engine'].astype(int)
train_data['Power'] = train_data['Power'].astype(float)
train_data['Mileage'] = train_data['Mileage'].astype(float)

Fuel = train_data['Fuel_Type']
Fuel = pd.get_dummies(Fuel, drop_first = True, dtype=int)

Trans = train_data['Transmission']
Trans = pd.get_dummies(Trans, drop_first = True, dtype=int)

data_train = pd.concat([train_data, Brand, Loc, Fuel, Trans], axis = 1)

data_train.drop(["Name", "Location",
"Fuel_Type",'Transmission','Price'], axis = 1, inplace = True)

#TEST DATA PREPROCESSING
dataframe_test = pd.read_csv('E:/test-data.csv')

test_data = dataframe_test.copy()
```

```
for i in range(0, len(test_data)):
    if test_data['Power'][i] == 'null bhp':
        test_data['Power'][i] = np.nan

for i in range(0, len(test_data)):
    if test_data['Mileage'][i] == '0.0 kmpl' or test_data['Mileage'][i] == '0.0 km/kg':
        test_data['Mileage'][i] = np.nan

for i in range(0, len(test_data)):
    if test_data['Engine'][i] == 'null CC' or test_data['Engine'][i] == '0 CC':
        test_data['Engine'][i] = np.nan

test_data.drop(['New_Price'], axis=1, inplace=True)
test_data.drop(['Unnamed: 0'], axis=1, inplace=True)
test_data.dropna(inplace = True)

test_data.reset_index(inplace = True)
```

```
test_data.drop(['index'], axis=1, inplace=True)

City_test = test_data['Location'].unique()

brand_test=[]

for i in range(0, 1195):
    k = test_data['Name'][i].split()
    brand_test.append(k[0].upper())

Brand_test = np.array(brand_test)

Brand_test = pd.get_dummies(Brand_test, drop_first = True,
dtype=int)

unique_brands_test=[]
for i in range(0,1195):
    if brand_test[i] in unique_brands_test:
        continue
    else:
```

```
        unique_brands_test.append(brand_test[i])

Loc_test = test_data['Location']

Loc_test = pd.get_dummies(Loc_test, drop_first = True, dtype=int)

test_data['Seats'] = test_data['Seats'].astype(int)

test_data.replace({'First': 1, 'Second': 2, 'Third': 3, 'Fourth & Above':
4}, inplace = True)

for i in range(0, 1195):
    k = test_data['Mileage'][i].split()
    test_data['Mileage'][i] = k[0]

for i in range(0, 1195):
    k = test_data['Power'][i].split()
    test_data['Power'][i] = k[0]

for i in range(0, 1195):
```

```
    k = test_data['Engine'][i].split()
    test_data['Engine'][i] = k[0]

test_data['Engine'] = test_data['Engine'].astype(int)
test_data['Power'] = test_data['Power'].astype(float)
test_data['Mileage'] = test_data['Mileage'].astype(float)

Fuel_test = test_data['Fuel_Type']
Fuel_test = pd.get_dummies(Fuel_test, drop_first = True, dtype=int)

Trans_test = test_data['Transmission']
Trans_test = pd.get_dummies(Trans_test, drop_first = True, dtype=int)

data_test = pd.concat([test_data, Brand_test, Loc_test, Fuel_test,
Trans_test], axis = 1)

data_test.drop(["Name", "Location", "Fuel_Type",'Transmission'],
axis = 1, inplace = True)

# FEATURE SCALING
```

```
#Not required for Random Forest Algorithm


#Feature Selection

X = data_train.copy()


plt.figure(figsize = (18,18))

sns.heatmap(train_data.corr(), annot = True, cmap = "RdYlGn")


from sklearn.ensemble import ExtraTreesRegressor

selection = ExtraTreesRegressor()

selection.fit(X, y)


plt.figure(figsize = (12,8))

feat_importances = pd.Series(selection.feature_importances_,
index=X.columns)

feat_importances.nlargest(20).plot(kind='barh')

plt.show()


#Fitting

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,
random_state = 42)

from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor()
regressor.fit(X_train, y_train)

y_pred = regressor.predict(X_test)

regressor.score(X_train, y_train)
regressor.score(X_test, y_test)

sns.distplot(y_test-y_pred)

plt.scatter(y_test, y_pred, alpha = 0.5)
plt.xlabel("y_test")
plt.ylabel("y_pred")
plt.show()

from sklearn import metrics
```

```
print('MAE:', metrics.mean_absolute_error(y_test, y_pred))

print('MSE:', metrics.mean_squared_error(y_test, y_pred))

print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))


from sklearn.model_selection import RandomizedSearchCV

#Randomized Search CV


# Number of trees in random forest

n_estimators = [int(x) for x in np.linspace(start = 100, stop = 1200,
num = 12)]

# Number of features to consider at every split

max_features = ['auto', 'sqrt']

# Maximum number of levels in tree

max_depth = [int(x) for x in np.linspace(5, 30, num = 6)]

# Minimum number of samples required to split a node

min_samples_split = [2, 5, 10, 15, 100]

# Minimum number of samples required at each leaf node

min_samples_leaf = [1, 2, 5, 10]


random_grid = {'n_estimators': n_estimators,
```

```
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf}

rf_random = RandomizedSearchCV(estimator = regressor,
param_distributions =
random_grid,scoring='neg_mean_squared_error', n_iter = 10, cv = 5,
verbose=2, random_state=42, n_jobs = 1)

rf_random.fit(X_train,y_train)

prediction = rf_random.predict(X_test)

plt.figure(figsize = (8,8))
sns.distplot(y_test-prediction)
plt.show()

plt.figure(figsize = (8,8))
plt.scatter(y_test, prediction, alpha = 0.5)
```

```
plt.xlabel("y_test")

plt.ylabel("y_pred")

plt.show()


print('MAE:', metrics.mean_absolute_error(y_test, prediction))

print('MSE:', metrics.mean_squared_error(y_test, prediction))

print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test,
prediction)))


rf_random.score(X_train, y_train)

rf_random.score(X_test, y_test)


#Save Model

import pickle

file = open('car.pkl', 'wb')


pickle.dump(regressor, file)

# -- coding: utf-8 --

"""
```

GitHub & Project Demo Link

https://github.com/IBM-EPBL/IBM-Project-48272-1660806218