KONGUNADU COLLEGEOFNGINEERINGANDTECHNOLOGY
(AUTONOMOUS)
Tholurpatti (P.O), Thottiam –T.K, Trichy – 621 215.
Department of Electronics and Communication Engineering

PROJECT TITLE – CAR RESALE VALUEPREDICTION

Domain of the Project : Applied Data Science

Batch ID : B12-6A2E

Team ID : PNT2022TMID13612

Academic Year : 2022-2023

Year/Semester : IV/VII

TEAM MEMBERS:

VEERAMANI S (621319106099)

RAGHUL G (621319106072)

SURAJ R (621319106093)

KARTHIKEYAN M (621319106310)

TEAM MENTOR:

DINESHKUMAR T

CONTENTS

S.NO	TITLE	PAGE NO
1.	**INTRODUCTION**	**4**
	1.1 Project Overview	4
	1.2 Purpose	6
2.	**LITERATURE SURVEY**	**7**
	2.1 Existing problem	11
	2.2 References	13
	2.3 Problem Statement Definition	15
3.	**IDEATION & PROPOSED SOLUTION**	**15**
	3.1 Empathy map canvas	15
	3.2 Ideation & Brainstorming	18
	3.3 Proposed solution	19
	3.4 Problem solution fit	22
4.	**REQUIREMENT ANALYSIS**	**23**
	4.1 Functional requirement	24
	4.2 Non-Functional requirement	25
5.	**PROJECT DESIGN**	**27**
	5.1 Data Flow Diagram	27
	5.2 Solution & Technical Architecture	28
	5.3 User Stories	30
6.	**PROJECT PANNING & SCHEDULING**	**31**
	6.1 Sprint Planning & Estimation	32
	6.2 Spring Delivery Schedule	33
	6.3 Reports From JIRA	35

7.	**CODING & SOLUTIONING**	**36**
	7.1 Feature 1	36
	7.2 Feature 2	41
8.	**TESTING**	**50**
	8.1 Test Cases	50
	8.2 User Acceptance Testing	51
9.	**RESULTS**	**53**
	9.1 Performance Metrics	53
10.	**ADVANTAGES AND DISADVANTAGES**	**54**
11.	**CONCLUSION**	**54**
12.	**FUTURE SCOPE**	**55**
13.	**APPENDIX**	**56**
	Source code	
	Github & Project Demo Link	

CHAPTER 1
INTRODUCTION

Today, the transportation industry is considered to be one of the backbones of the economy. Automobiles are referred to as the "Industry of Industries" in developed nations. According to industry professionals, the UAE's automotive industry has seen remarkable growth. Besides being the fastest-growing nation in the automobile industry, it represents its global presence. In Dubai, like most other countries, cars are gaining a great deal of popularity among the local population and the ex-pat community who work in the country. There are used cars for sale in the UAE of all makes and models, even cars from well-known brands (Rizvi, 2019). UAE's auto industry is experiencing constant growth, registered at 27%, with a total industry volume (TIV) of 310,403 cars. Approximately 1.49 million units were sold within the Gulf Cooperation Council .

1.1PROJECT OVERVIEW

Cooperation Council countries are growing at 10% in 2021 (Research, 2020). So far, the market in the UAE has grown by 19%. It is thus the world's largest market in terms of growth rate. Almost everyone wants their own car these days, but because of factors like affordability or economic conditions, many prefer to opt for pre-owned cars. Accurately predicting used car prices requires expert knowledge due to the nature of their dependence on a variety of factors and features. Used car prices are not constant in the market, both buyers and sellers need an intelligent system that will allow them to predict the correct price efficiently. In this intelligent system, the most difficult problem is the collection of the dataset which contains all important elements like the manufacturing year of the car, its gas type, its condition, miles

driven, horsepower, doors, number of times a car has been painted, customer reviews, the weight of the car, etc. It is clear that the price of the product is affected by many factors, but unfortunately, information about these features is not always readily available. Since this project primarily focuses on the Dubai market, the benchmark dataset containing all key features is scraped. It is necessary to pre-process and transform collected data in the proper format prior to feeding it directly to the data mining model. As a first step, the dataset was statistically analyzed and plotted. Missing, duplicated, and null values were identified and dealt with. Features were chosen and extracted using 2 correlation matrices. To build an efficient model, the most correlated features were retained, and others were discarded. This prediction problem can be considered a regression problem since it belongs to the supervised learning domain. Three Regressor known as random forest, linear regression, and bagging regression were trained and compared. A random forest Regressor outperformed all others in this project, so it was chosen as the main algorithm model.

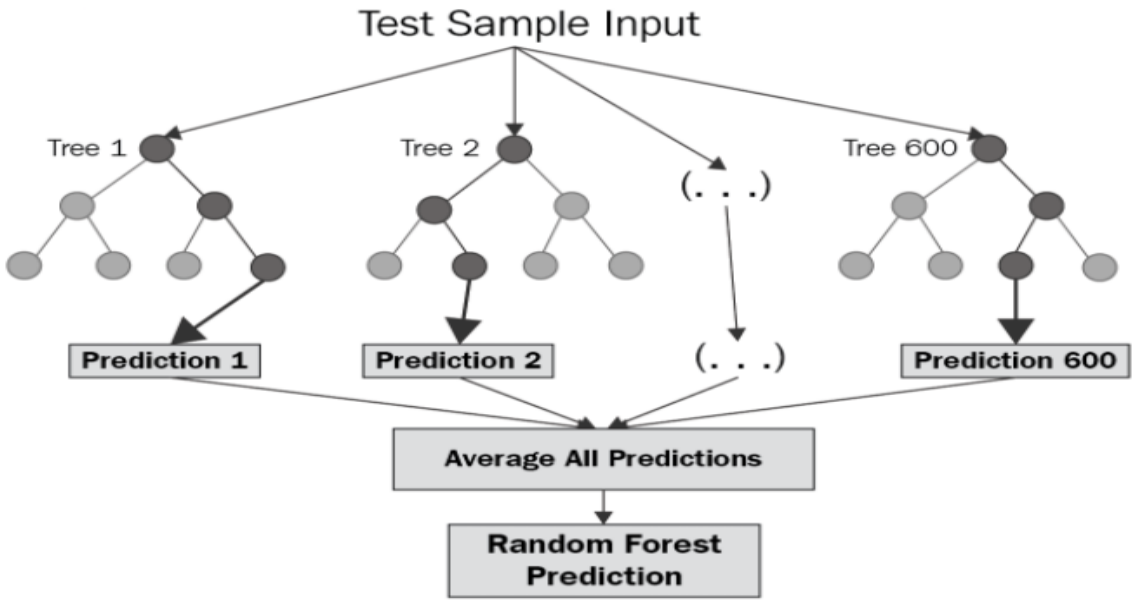


Fig 1.1-PROJECT FLOW

Above image tells that the –

Three Regressor known as random forest, linear regression, and bagging regression were trained and compared.

1.2 PURPOSE

The system is defined in the python language that predicts the amount of resale value based on the given information. The system works on the trained dataset of the machine learning program that evaluates the precise value of the car. User can enter details only of fields like purchase price of car, kilometers driven, fuel of car, year of purchase. Up on form submission, the data is sent to the ML model via Flask API and the model responds with a predicted resale value of the car based on user input. This prediction is displayed on the web page using a render template. Thus, with minimal information and without human intervention or manual examination, a user can predict the resale value of his car.We have used web scraping libraries to gather data from the webpages of cars24 website. The script runs and captures data from the HTML div mentioned in the code via URL. URL should be entered by the user. For now, we have captured data by entering URL for Swift Dzire cars for 5 cities. The second part is the web-based car resale value prediction. We have trained a boosting algorithm-based ML model using data from the previous step after preprocessing and cleaning. The trained model is used for prediction. The front-end form asks users to fill values which are required for the ML model to make prediction IE- city, kms driven, year of purchase and fuel type. Upon form submission, the data is sent to the ML model via Flask API and the model responds with a predicted resale value of the car based on user input. This prediction is displayed on the web page

using a render template. Thus, with minimal information and without human intervention or manual examination, a user can predict the resale value of his car.

CHAPTER 2

LITERATURE SURVEY

Price Prediction of Used Cars Using MachineLearning.

1. Chuyang Jin& 2021IEEE International Conference on Emergency Science and Information Technology (ICESIT)

This paper aims to build a model to predict used cars' reasonable prices based on multiple aspects, including vehicle mileage, year of manufacturing, fuel consumption, transmission, road tax, fuel type, and engine size. This model can benefit sellers, buyers, and car manufacturers in the used cars market. Upon completion, it can output a relatively accurate price prediction based on the information that users input. The model building process involves machine learning and data science. The dataset used was scraped from listings of used cars. Various regression methods, including linear regression, polynomial regression, support vector regression, decision tree regression, and random forest regression, were applied in the research to achieve the highest accuracy. Before the actual start of model-building, this project visualized the data to understand the dataset better. The dataset was divided and modified to fit the regression, thus ensure the performance of the regression. To evaluate the performance of each regression, R-square was calculated. Among all regressions in this project, random forest achieved the highest R-square of 0.90416. Compared to previous research, the resulting model includes more aspects of used cars while also having a higher prediction accuracy.

2. Prediction Of Used Car Prices Using Artificial Neural Networks And Machine Learning C. Lakshmi & 2022 International Conference on Computer Communication and Informatics (ICCCI)

With the extensive growth in usage of cars, the newly produced cars are unable to reach the customers for various reasons like high prices, less availability, financial incapability, and so on. Hence the used car market is escalated across the globe but in India, the used car market is in a very nascent stage and mostly dominated by the unorganized sector. This gives chance for fraud while buying a used car. Hence a high precision model is required which will estimate the price of an used car with none bias towards customer or merchandiser.In this model, A Supervised learning-based Artificial Neural Network model and Random Forest Machine Learning model are developed which can learn from the car dataset provided to it. This project presents a working model for used car price prediction with a low error value. A considerable number of distinct attributes are examined for reliable and accurate predictions. The results obtained agree with theoretical predictions and have shown improvement over models which use simple linear models. An ANN (Artificial Neural Network) is built by using Keras Regression algorithm namely Keras Regressor and other Machine Learning Algorithms namely Random Forest, Lasso, Ridge, Linear regressions are built. These algorithms are tested with the car dataset. Experimental results have shown that the Random Forest model with a Mean Absolute Error value of 1.0970472 and R2 error value of 0.772584 has given the less error among all the other algorithms. The work presented here has shown profound implications for future studies of Used Cars price Prediction using Random Forest and might one day help to solve the problem of fraudswith one hundred percent accuracy.

3. Used Car Price Prediction using Machine Learning Abderrahim Beni-Hssane & 2022 11th International Symposium on Signal, Image, Video and Communications (ISIVC)

In many business fields that are related to statistics and machine learning (ML), multiple linear regression (MLR) models are often used to estimate and fit a linear relationship between a continuous response variable and other explanatory variables. In our case study, we applied several regression techniques based on supervised machine learning to predict the resale price of used cars given many factors such as mileage, fuel type, fiscal power, mark, model, and the production year of the car. In all tested models, gradient boosting regressor showed a high R-squared score and low root mean square err

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

Feng Wang & 2021 International Conference on Networking, Communications and Information Technology (NetCIT)

In this paper, we use machine learning algorithms to predict the price of used cars with less human intervention to make the results more objective. The method used is to preprocess the dataset through Python's Pycaret package and compare the performance of each algorithm through the algorithm comparison function, in this study Extra Trees Regressor, Random Forest Regressor performs relatively well. Finally, the algorithm was optimized by using the hyperparameter function. The results show that R2 = 0.9807 obtained from extreme random numbers is the best performance. The algorithm was obtained and validated with new data to derive the final algorithm model. When new used car data flows into the used car system, used car prices will be automatically generated by this algorithm, which will make the

workflow of the used car market faster and more competitive for that used car market.

5. An Automated Car Price Prediction System Using Effective Machine Learning Techniques Santosh Kumar Satapathy & 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)

This research focuses on Building a mathematical model that could predict the price of a second-hand car based on its current features. Determining the price of a used automobile is a difficult task because several factors like Current Mileage, Current Condition, Make, Year, etc., can influence the prediction prices of an automobile. And, from the perspective of a person who sells, it becomes a dilemma to predict the price of a second-hand car accurately. Thus, the point of interest of this challenge is in growing gadgets, studying models that can correctly expect the price of a used car primarily based on its capabilities. Due to this, in turn, a consumer can make a much more informed purchase. Therefore, We will be implementing and examining various Machine Learning Techniques with Data Analysis to Provide an Accurate and Easy to use solution.

2.1 EXISTING PROBLEM

As indicated by author Sameer Chand, they have done the forecasts of vehicle cost from the chronicled information that has been gathered from every day papers. They have utilized the administered AI strategies for foreseeing the cost of vehicles. Numerous different calculations like various straight relapse, k-closest neighbor calculations, gullible based, and some choice tree calculations additionally been utilized. Every one of the four calculations are looked at and tracked down the best calculation for forecast. They have confronted a few challenges in looking at the calculations, by one way or another they have overseen. As indicated by creators Pattabiraman, this paper is more focused on the connection among vender and purchaser. To foresee the cost of four wheelers, more highlights are required like previously given value, mileage, make, model, trim, type, chamber, liter, entryways, voyage, sound, cowhide. Utilizing these highlights the cost of vehicle has been anticipated with the assistance of factual investigation framework for exploratory information examination. As per creators EnisGegic et al, in this paper the chiefly focus on gathering different information from web entryway by utilizing web scrap methods. Furthermore, those have been contrasted and the assistance of various AI calculations to foresee the vehicle cost in simple way. They arranged the value as per various scopes of value that is as of now given. Fake neural organization, support vector machine, arbitrary timberland calculations were utilized on various datasets to construct classifiers model. Another methodology was given by Richardson in his postulation work. In his hypothesis it states more strong vehicles will be delivered by vehicle maker. He looked at the crossover vehicles and conventional vehicles in scraper it really holds their incentive for longer time utilizing numerous relapse procedures. This works on the natural conditions, and furthermore it assists with giving colossal effectiveness of utilizing energizes. Wu et al, in this paper they have

utilized neuro fluffy informationbased framework to exhibit vehicle value forecast. By considering the accompanying ascribes like brand, year of creation and sort of motor they anticipated a model which has comparative outcomes as the basic relapse model. Additionally, they made a specialist framework named ODAV (Optimal Distribution of Auction Vehicles) as there is a popularity for selling the by vehicles toward the finish of the renting year by vehicle vendors. This framework gives experiences into the best costs for vehicles, just as the area where all that cost can be acquired.

2.2 REFERENCES

1. Adam A, Mokhtar N, Mubin M, Ibrahim Z, Tumari MZ, Shapiai MI (2014), 'Feature selection and classifier parameter estimation for EEG signal peak detection using gravitational search algorithm', 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, (pp. 103-108).
2. Akash Gupta, Rohini Srivastava, Himanshu Gupta and Basant Kumar (2020), 'IoT Based Fall Detection Monitoring and Alarm System For Elderly', on IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON).
3. Alshurafa N, Eastwood JA, Nyamathi S, Liu JJ, Xu W, Ghasemzadeh H, Pourhomayoun M, Sarrafzadeh M (2014), 'Improving compliance in remote healthcare systems through smartphone battery optimization', IEEE Journal of Biomedical and Health Informatics, 19(1):57-63.
4. AmberlayRuíz-Serrano; Miriam C. Reyes-Fernández; Rubén Posada-Gómez; Albino Martínez-Sibaja; Alberto A. Aguilar-Lasserre(2014), 'Obstacle avoidance embedded system for a smart wheelchair with a multimodal navigation interface', on 11th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE).
5. Badal S, Sai Ravela, Bruce A. Draper, Allen Hanson (1995), 'A practical obstacle detection and avoidance system' on Conference on Applications of Computer Vision, 10.1109/ACV.1994.341294.

6. Bianchi F, Redmond SJ, Narayanan MR, Cerutti S, Lovell NH (2010), 'Barometric pressure and triaxial accelerometry-based falls event detection', IEEE Transactions on Neural Systems and Rehabilitation Engineering, 18(6):619-27.

7. Changhong Wang, Michael R Narayanan, Stephen R Lord, Stephen J Redmond, Nigel H Lovell (2014), 'A low-power fall detection algorithm based on triaxial acceleration and barometric pressure', Annu Int Conf IEEE Eng Med BiolSoc, doi: 10.1109/EMBC.2014.6943655.
8. Devansh Kumar Garg and Gauri Rao (2020), 'An IoT Based Fall Detection System', on International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-9 Issue-6.
9. Dr.EmnaBaklouti, Prof. Nader Ben Amor & Prof. Mohammed Jallouli(2016), 'Autonomous wheelchair navigation with real time obstacle detection using 3D sensor', on Journal for Control, Measurement, Electronics, Computing and Communications.
10. Fabio Bagalà, Clemens Becker, Angelo Cappello, Lorenzo Chiari, KamiarAminian, Jeffrey M Hausdorff, WiebrenZijlstra, Jochen Klenk(2012), 'Evaluation of accelerometer-based fall detection algorithms on real-world falls',7(5):e37062. doi: 10.1371/journal.pone.0037062.

.

2.3 PROBLEM STATEMENT DEFINITION

Machine Learning has become a tool used in almost every task that requires estimation. Companies like Cars24 and Cardeho.com uses Regression analysis to estimate the used car prices. So we need to build a model to estimate the price of used cars. The model should take car-related parameters and output a selling price. The selling price of a used car depends on certain features as mentioned below:

- Fuel Type
- Manufacturing Year
- Miles Driven
- Number of Historical Owners
- Maintenance Record

This is a supervised learning problem and can be solved using regression techniques. We need to predict the selling price of a car based on the given car's features. Supervised Regression problem require labeled data where our target or dependent variable is the selling price of a car. All other features are independent variables.

Following are some regression algorithms that can be used for predicting the selling price.

- Liner Regression
- Decision Tree Regressor
- Random Forest Regressor

Linear Models are relatively less complex and explainable, but linear models perform poorly on the containing the outliers. Linear models fail to perform well on nonlinear datasets. In such cases, non-linear regression algorithm Random Forest

Regressor and XGBoost Regressor perform better in fitting the nonlinear data. In this, will use Random Forest Regressor for predicting the selling price of cars. Our data contains some outliers, and treating them is entirely possible, but the performance of nonlinear regression models is insensitive to outliers.

CHAPTER 3

IDEATION & PROPOSED SOLUTION

Ideation is the process where you generate ideas and solutions through sessions such as Sketching, Prototypeing, Brainstorming, Brainwriting, Worst Possible Idea, and a wealth of other ideation techniques. Ideation is also the third stage in the Design Thinking process. Ideation is often the most exciting stage in a Design Thinking project, because during Ideation, the aim is to generate a large quantity of ideas that the team can then filter and cut down into the best, most practical or most innovative ones in order to inspire new and better design solutions and products.

3.1 EMPATHY MAP CANVAS

An empathy map is a collaborative tool teams can use to gain a deeper insight into their customers. Much like a user persona, an empathy map can represent a group of users, such as a customer segment. The empathy may was originally created by Dave Gray and has gained much popularity within the agile community.

Everyone would add at least one sticky to every section. You might ask questions,

such as:

- What would the user see while using our product in their environment?
- What might the user be saying and/or doing while using our product? How

would that change in a public or private setting?

- What are some of the user's pain points or fears when using our product?
- What gains might the user experience when using our product?

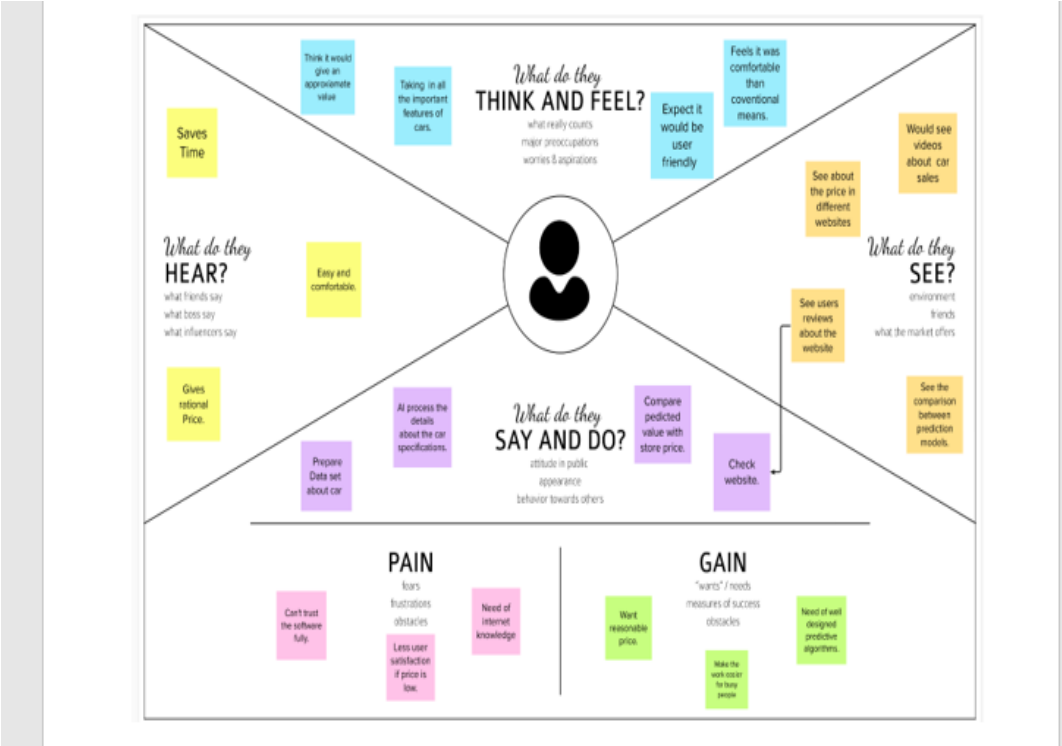
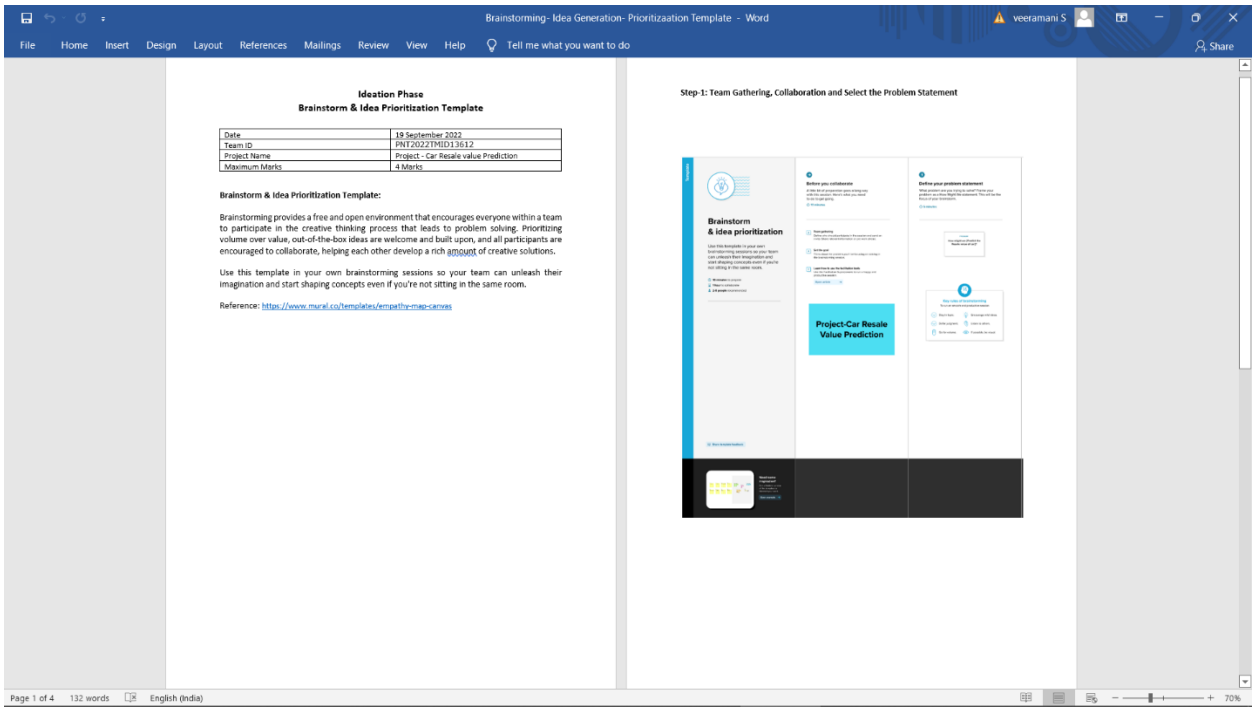


Fig.3.1.1 EMPATHY MAP CANVAS

3.2 IDEATION & BRAINSTORMING

A brainstroming session involves people meeting(either in person or virtually), to discuss, draw out, explain, and write down ideas to slove problems or to create something new.

For example, a major corporation that recently learned it is the object of a major lawsuit may want to gather together top executives for a brainstorming session on how to publicly respond to the lawsuit being filed.Participants in a brainstorming session are encouraged to freely toss out whatever ideas may occur to them. The thinking is that by generating a large number of ideas, the brainstorming group is likely to come up with a suitable solution for whatever issue they are addressing.The lines between ideation and brainstorming have become a bit more blurred with the development of several brainstorming software programs, such as Bright idea and Idea wake. These software programs are designed to encourage employees of companies to generate new ideas for improving the companies' operations and,ultimately, bottom-line profitability.



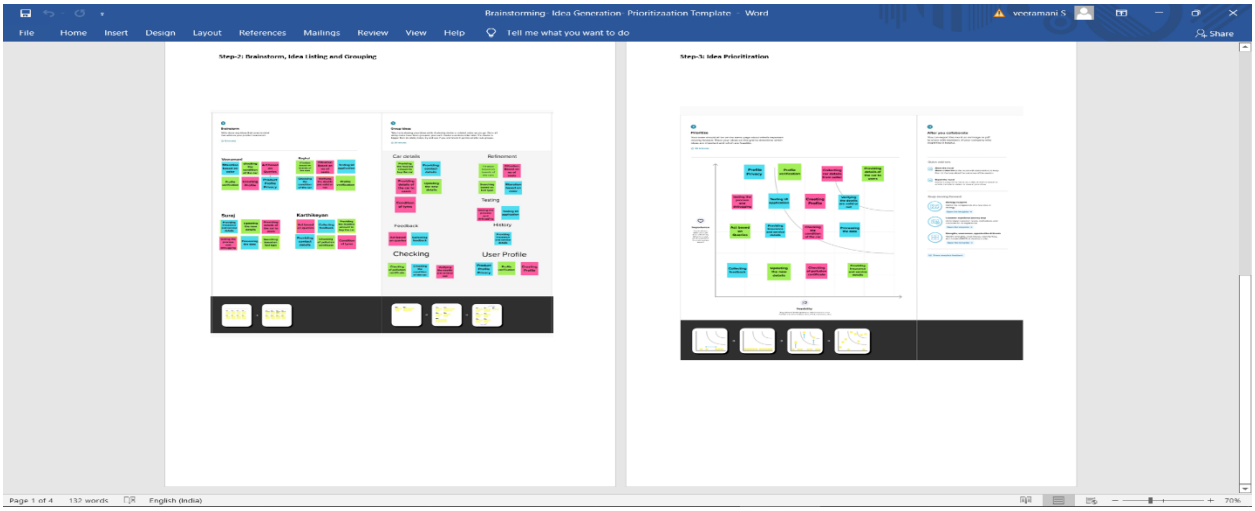


Fig.3.2.1 IDEATION & BRAINSTORMING

3.3 PROPOSED SOLUTION

1. PROBLEM STATEMENT (PROBLEM TO BE SOLVED)

To build a prediction system which can predict the price of the second-hand used cars.

2. IDEA/SOLUTION DESCRIPTION

To foresee the resale worth of the vehicle, we proposed a keen, adaptable, and powerful framework that depends on utilizing relapse calculations. Leading organizations are collecting tons of data every day to derive business decisions and solutions from it. With this huge amount of data, demand for data scientists and data analysts is massively increasing. Nowadays, Machine Learning and Artificial Intelligence are applicable in almost every sector. Companies are adopting smart AI solutions in their product to eliminate manual interventions.

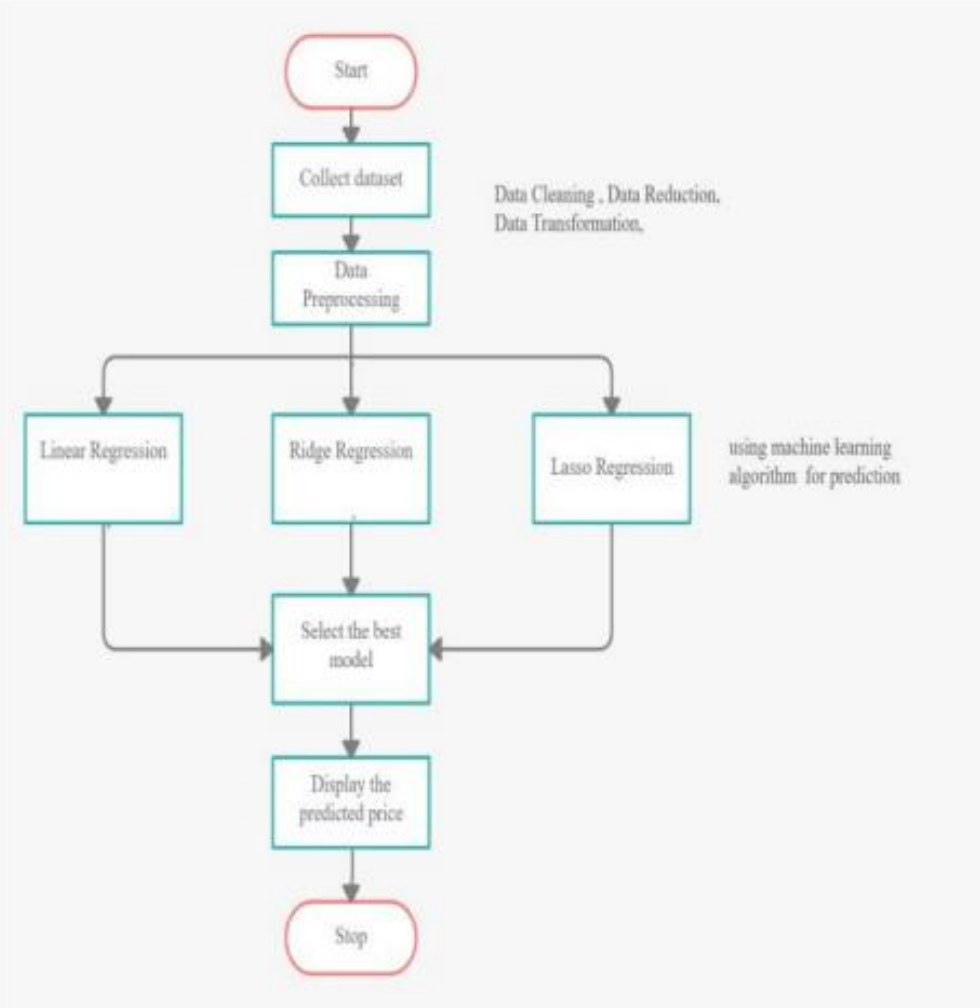


Fig 3.3- SOLUTION FLOW

The above figure tells that the with this huge amount of data, demand for data scientists and data analysts is massively increasing. Nowadays, Machine Learning and Artificial Intelligence are applicable in almost every sector.

3. NOVELTY/UNIQUENESS

The dealers of used cars have instructive benefits over statistical surveying organizations, which empower them to gauge resale costs all the more precisely. This suggests that merchants have a motivating force to put resources into in-house determining arrangements. Rather than putting together their evaluating choices with respect to remotely created lingering esteem gauges.

4. SOCIAL IMPACT/CUSTOMER SATISFACTION

This is a supervised learning problem and can be solved using regression techniques. We need to predict the selling price of a car based on the given car's features.

5. BUSINESS MODEL (FINANCIAL BENEFIT)

The rate of insurance is generally based on the age of the car. Newer the vehicle, higher is the cost towards insurance and vice versa. The rate of insurance for pre-owned vehicles is therefore lower. Besides insurance, you also have to pay a lower amount towards registration fees. As the cost of registration is based on the transaction price of the car, buying a used car reduces the cost of registration. As used cars come with a lower price tag, the amount you may have to borrow will be lower. Many financial institutes offer used car loans with higher borrowing amount and attractive interest rates.

6. SCALABILITY OF SOLUTION

The pre-owned car market is witnessing a growth trajectory like never before. It is acutely seizing the opportunity to capitalize on the revelations happening in the automobile sector to turn the wind in their favor. Currently where India calibrates a pre-owned car market size 1.5 times more than the new car sales, it is estimated that in near future it will outpace the new car market.

3.4 PROBLEM SOLUTION FIT

Problem-solution fit is a term used to describe the point validating that the base problem resulting in a business idea really exists and the proposed solution actually solves that problem. Validate that the problem exists, when you validate your problem hypothesis using real-world data and feedback.

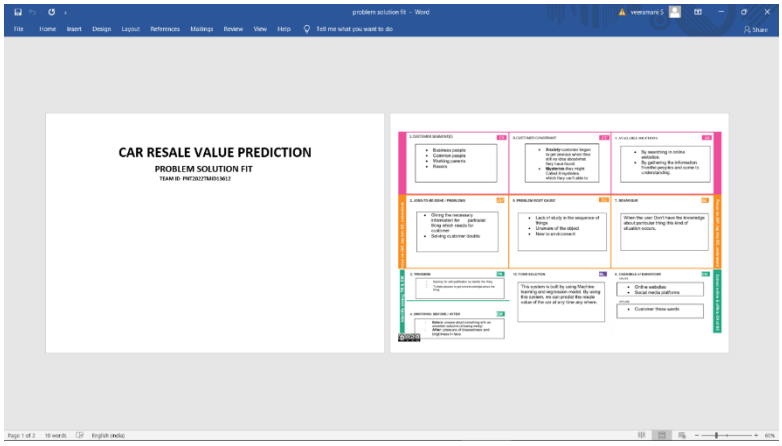


Fig.3.4.1- Problem solution fit

The above diagram show that the term used to the point validating that the base problem resulting in a business idea really exists and the proposed solution .

CHAPTER 4

REQUIREMENT ANALYSIS

Requirements analysis, also called requirements engineering, is the process of determining user expectations for a new or modified product. These features, called requirements, must be quantifiable, relevant and detailed. In software engineering, such requirements are often called functional specifications. Requirements analysis is an important aspect of project management.

Requirements analysis involves frequent communication with system users to determine specific feature expectations, resolution of conflict or ambiguity in requirements as demanded by the various users or groups of users, avoidance of feature creep and documentation of all aspects of the project development process from start to finish. Energy should be directed towards ensuring that the final system or product conforms to client needs rather than attempting to mold user expectations to fit the requirements. Requirements analysis is a team effort that demands a combination of hardware, software and human factors engineering expertise as well as skills in dealing with people. The purpose of the Requirements Analysis Phase is to transform the needs and high-level requirements specified in earlier phases into unambiguous (measurable and testable), traceable, complete, consistent, and stakeholder-approved requirements.

4.1 FUNCTIONAL REQUIREMENT

Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a

system is supposed to accomplish. Behavioural requirements describe all the cases where the system uses the functional requirements, these are captured in use cases.

Following are the functional requirements of the proposed solution.

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	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Login	The login screen is used to verify the identity of the user. The account can be accessed using the user's registered email address and password.
FR-4	Categories	On the main page, we can see the categories of second-car models that are fall into the given list of parameters.
FR-5	Enter the parameters like car model, year of manufacturing, mileage, fuel type, engine type	Retrieve the data related to the given parameters from the dataset.

4.2 NON-FUNCTIONAL REQUIREMENTS

In systems engineering and requirements engineering, a non-functional requirement (NFR) is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviours. They are contrasted with functional requirements that define specific behaviour or functions. The plan for implementing non-functional requirements is detailed in the system architecture the system architecture, because they are usually architecturally significant requirements.

Following are the non-functional requirements of the proposed solution.

FR No.	NON-FUNCTIONAL REQUIREMENT	DESCRIPTION
NFR-1	Usability	This resale value prediction system is made for general purpose to just predict the amount that can be roughly acquired by the user.
NFR-2	Security	A security requirement is a statement of needed security functionality that ensures one of many different security properties of software is being satisfied.
NFR-3	Reliability	The system has to be 100% reliable due to the importance of data and the damages that can be caused by incorrect or incomplete data. The system will run 7 days a week. 24 hours a day.

NFR-4	Performance	The information is refreshed depending upon whether some updates have occurred or not in the application. The system shall respond to the member in not less than two seconds from the time of the request submittal. The system shall be allowed to take more time when doing large processing jobs. Responses to view information shall take no longer than 5 seconds to appear on the screen.
NFR-5	Availability	The system is available 100% for the user and is used 24 hrs a day and 365 days a year. The system shall be operational 24 hours a day and 7 days a week.
NFR-6	Scalability	Scalability is the measure of a system's ability to increase or decrease in performance and cost in response to changes in application and system processing demands.

CHAPTER 5

PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

A data flow diagram shows the way information flows through a process or system. It includes data inputs and outputs, data stores, and the various subprocesses the data move through. DFDs are built using standardized symbols and notation to describe various entities and their relationships

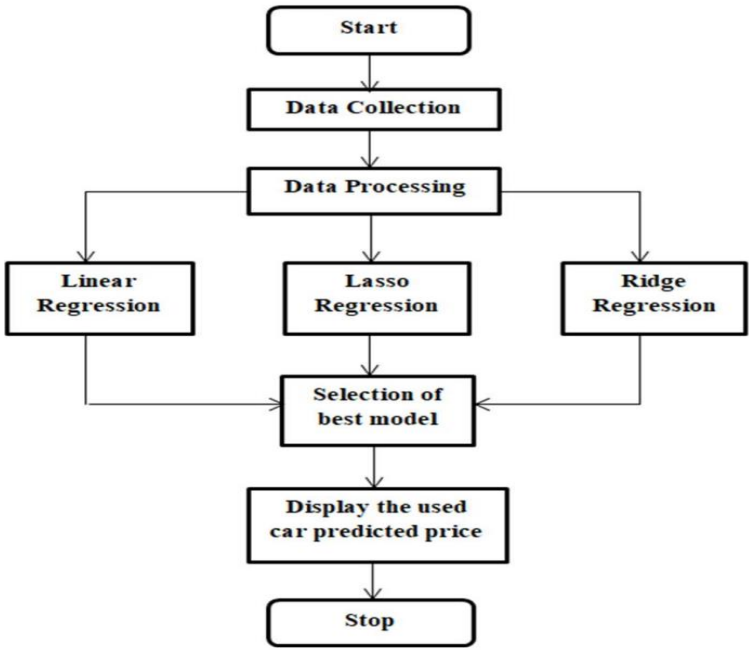


Fig.5.1.1 DATA FLOW DIAGRAMS

The above diagram shows that the includes data inputs and outputs, data stores, and the various subprocesses the data move through. D

5.2 SOLUTION & TECHNICAL ARCHITECTURE

Based on the complexity of the deployment, a solution architecture diagram may actually be a set of diagrams documenting various levels of the architecture. The diagram relates the information that you gather on the environment to both physical and logical choices for your architecture in an easily understood manner.

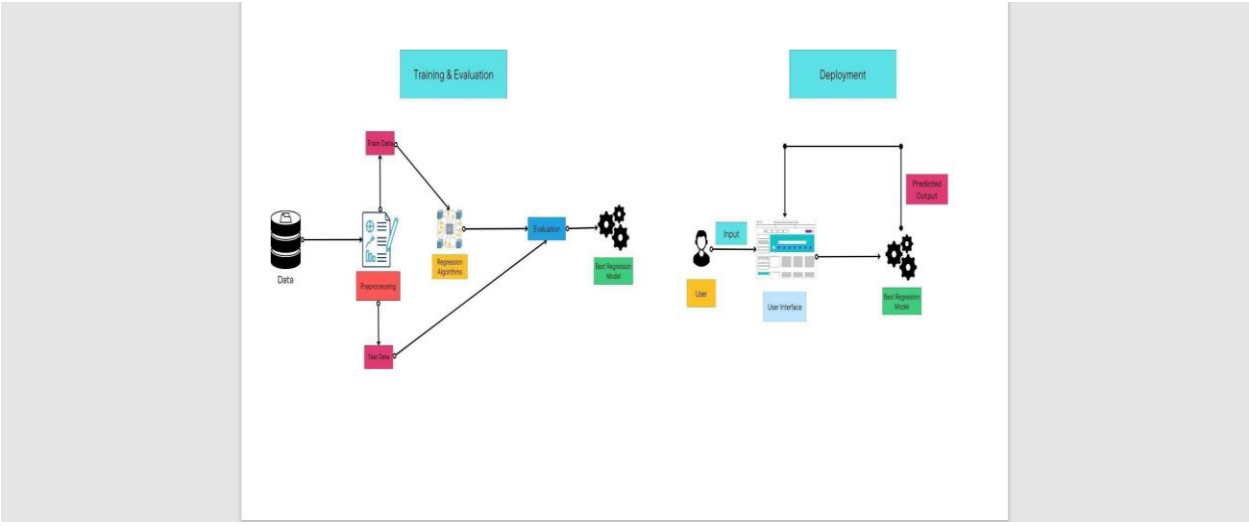


Fig.5.2.1 TECHNOLOGY STACK

The above diagram relates the information that you gather on the environment to both physical and logical choices for your architecture in an easily understood manner.

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	Technology of Opensource framework, Google collab, anaconda navigator, flask framework
2.	Security Implementations	List all the security / access controls implemented, use of firewalls etc.	Encryptions, IAM Controls, OWASP etc.
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	Google collab
4.	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)	Web application to access the system
5.	Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.	Linear regression

5.3 USER STORIES

A user story is the smallest unit of work in an agile framework. It's an end goal, not a feature, expressed from the software user's perspective. A user story is an informal, general explanation of a software feature written from the perspective of the end user or customer. the purpose of a user story is to articulate how a piece of work will deliver a particular value back to the customer. Note that customers don' t have to be external end users in the traditional sense, they can also be internal customers or colleagues within your organization who depend on your team. User stories are a few sentences in simple language that outline the desired outcome. They go into detail. Requirements are added later, once agreed upon by the team.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile 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 Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-1	As a user, I can see my dashboard and go through the functions provide by the system	I can access my dashboard	High	Sprint-1
Customer (Web user)	Registration		As a user, I can register for my account through web and login to my webpage			
Customer Care Executive	Login	USN-1	Make a call to the customer care executive and rectify the queries	Help the user to access the system	High	Sprint-1

CHAPTER 6

PROJECT PLANNING & SCHEDULING

Planning - Planning pertains to the process of creating a plan of which materials and resources will be required to fulfil incoming and forecasted demand. This step is crucial to ensure that you have enough materials and resource capacity available to produce your orders on time. This component pertains to the 'what' and 'how' of any project: what exactly needs to be achieved and how it will be accomplished.

Scheduling - Scheduling pertains to establishing the timing of the use of specific resources of that organization. In production, scheduling involves developing schedules for workers, equipment, and materials. It reflects on the 'when' of a project, by assigning the appropriate resources to get the production plan completed within a period of time. Creating optimized production schedules ensures that your facility is able to reduce costs, increase productivity, and deliver goods to customers on time.

In order to create accurate and realistic production plans that allow manufacturers to react quickly to changes, it is important to have a production plan that is aligned with the resource and material scheduling process. Having any discrepancy or divergence between the planning and scheduling process creates inefficiencies that can be costly for your business. The bigger the divergence, the larger the cost.

6.1 SPRINT PLANNING & ESTIMATION

PLANNING:

In Sprint Planning, the team decides what it will build in the upcoming Sprint and how they will build it. The team commits to the Sprint goal after breaking down user stories into tasks and doing task-level estimation. Sprint Planning is done by the

Product Owner, Scrum Master, and the Team. In Scrum, every project is broken into time blocks called sprints, usually 2-4 weeks long. A sprint planning meeting is done.

ESTIMATION

In Scrum Projects, Estimation is done by the entire team during Sprint Planning Meeting. The objective of the Estimation would be to consider the User Stories for the Sprint by Priority and by the Ability of the team to deliver during the Time Box of the Sprint.

Product Owner ensures that the prioritized User Stories are clear, can be subjected to estimation, and they are brought to the beginning of the Product Backlog.. The size of the Product Increment is estimated in terms of User Story Points. Once the size is determined, the effort is estimated by means of the past data, i.e., effort per User Story Point called Productivity.

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 register for the application by entering my email, password, and confirming my password.	2	High	Swetha S, Swetha S
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Pavithra J, Yuvarani S
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	Low	Swetha S, Swetha S
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	Pavithra J, Yuvarani S

Fig.6.1.1 SPRINT PLANNING & ESTIMATION

6.2 SPRINT DELIVERY SCHEDULE

Since sprints take place over a fixed period of time, it's critical to avoid wasting time during planning and development. And this is precisely where sprint scheduling enters the equation. In case you're unfamiliar, a sprint schedule is a document that outlines sprint planning from end to end. It's one of the first steps in the agile sprint planning process—and something that requires adequate research,

Project Tracker, Velocity & Burndown Chart: (4 Marks)

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

Velocity:

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

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

6.3 REPORTS FROM JIRA

A burn down 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. Typically, in a burn down chart, the outstanding work is often on the vertical axis, with time along the horizontal. It is useful for predicting when all of the work will be completed. In the Daily Scrum the Development Team updates the Sprint Burn Down and plots the remaining work of the day.

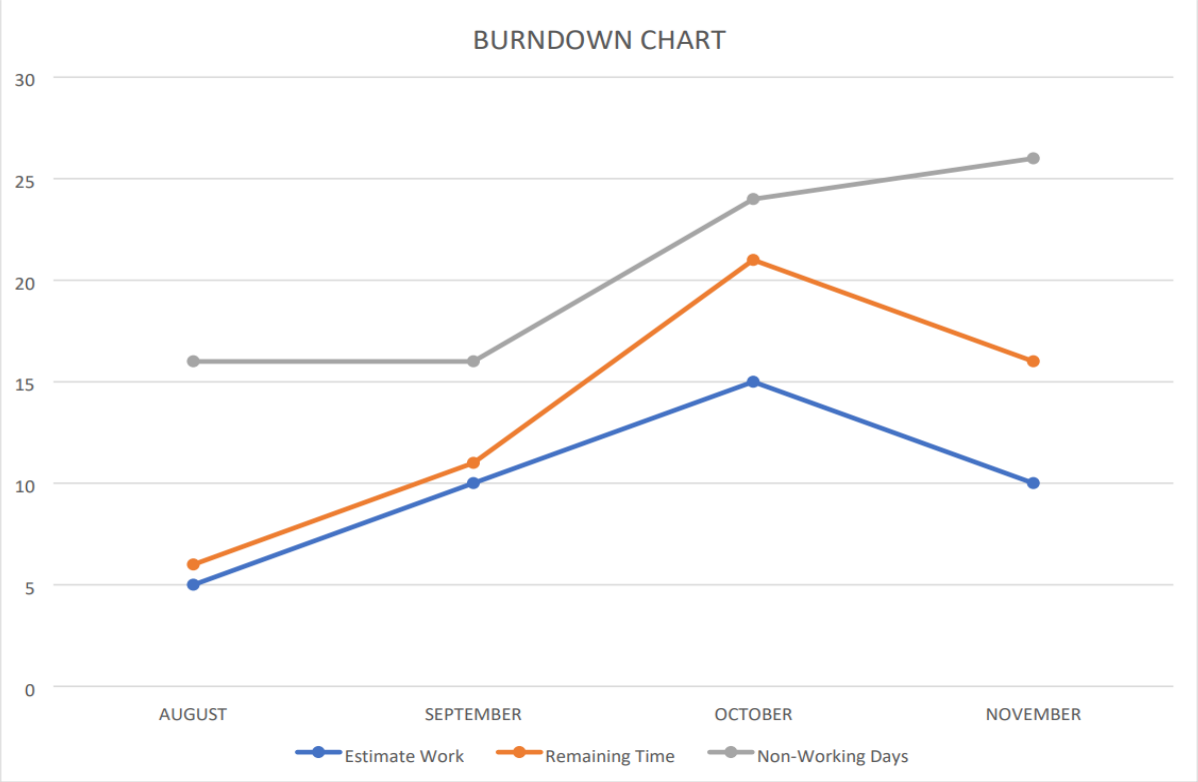


Fig6.3.1 REPORT FROM JIRA

The above diagram show the flow chart of the report from jira

CHAPTER 7

CODING AND SOLUTIONING

7.1 FEATURE 1

import pandas as pd

import numpy as np

df = pd.read_csv("car data.csv")

df.sample(5)

df.shape

let's first check na value

df.isna().values.any()

print("Fuel Type: ", df.Fuel_Type.unique())

print("Seller Type: ", df.Seller_Type.unique())

print("Transmission: ", df.Transmission.unique())

print("Owner: ", df.Owner.unique())

year indicates a purchased year of car, it's basically use for calculating how many years old thatcar

for that we need to subtract year from current year

from datetime import datetime

df["Current_year"] = datetime.now().year

now let's subtract Year from Current Year

df["Year"] = df.Current_year - df.Year

now let's drop current year we don't need it

df.drop(columns="Current_year", axis = 1, inplace = True)

let's encode text or categorical data using one hot encoding

let's create dummy variables for Fuel_Type Seller_Type and Transmission

dummy = pd.get_dummies(df[["Fuel_Type", "Seller_Type", "Transmission"]], drop_first =

```
True)
# drop first columns for preventing dummy variable trape
# for Fuel_Type CNG will be Removed, for Seller Type Dealer will be removed and fo
Transmission Automatic remove
df = pd.concat( [df, dummy], axis = 1)
# now we don&#39;t need that original text data columns so let&#39;s drop it
df.drop(columns = [&quot;Fuel_Type&quot;, &quot;Seller_Type&quot;, &quot;Transmission&quot;], inplace = True)
# let&#39;s get a corelation of our data
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
plt.figure(figsize = (10, 6))
sns.heatmap(df.corr(), annot = True)
plt.show()
# let&#39;s encode Car_Name text data using one hot encoding
# for this let&#39;s convert first Car_Names data into lowercase
df.Car_Name = df.Car_Name.str.lower()
dummy = pd.get_dummies(df.Car_Name, drop_first = True)
# here we dropped first columns which is 800 for prevent dummy variable trap
df = pd.concat([df, dummy], axis = 1)
# now we don&#39;t need Car_Name columns so let&#39;s drop it
df.drop(&quot;Car_Name&quot;, axis = 1, inplace = True)
# now let&#39;s create a feature matrix X and target vector y
X = df.drop(columns=&quot;Selling_Price&quot;)
y = df.Selling_Price
# let&#39;s find important feature using ExtraTreesRegressor model
from sklearn.ensemble import ExtraTreesRegressor
etr = ExtraTreesRegressor()
etr.fit(X, y)
```

important_features = etr.feature_importances_

let's get top important features

important_features = pd.Series(important_features, index = X.columns).sort_values(ascending = False)

let's plot top 5 important feature

plt.figure(figsize = (10, 6))

important_features[:5].plot(kind = "barh")

plt.title("Top 5 Important Feature")

plt.show()

let's divide our data into train and test part

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20)

so let's hypertuning parameter and find best algorithm with it's best parameter

from sklearn.linear_model import LinearRegression

from sklearn.tree import DecisionTreeRegressor

from sklearn.ensemble import RandomForestRegressor

let's create a dict for paramer

algos = {

"Linear Regression": {

"model": LinearRegression(),

"param":{

"normalize":[True, False]

}

},

"Decision Tree": {

"model": DecisionTreeRegressor(),

"param": {

"criterion": ["mse", "friedman_mse"],

"splitter": ["best", "random"]

}

},

"Random Forest": {

"model": RandomForestRegressor(),

"param": {

"n_estimators": [int(x) for x in np.linspace(100, 1200, 12)],

"max_features": ["auto", "sqrt"],

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

"min_samples_split": [2, 5, 10, 15, 100],

"min_samples_leaf": [1, 2, 5, 10]

}

}

}

let's find best algotithm with it's best parameter

here we are going to use Randomize Search cv for hyperparamter tuninig

from sklearn.model_selection import RandomizedSearchCV

best_models = {}

scores = []

for model_name, values in algos.items():

model_tunning = RandomizedSearchCV(values["model"], values["param"], n_iter=10, cv = 5,

n_jobs = -1)

model_tunning.fit(X_train, y_train)

best_models[model_name] = model_tunning

scores.append({

"Model": model_name,

"Best Parameters": model_tunning.best_params_,

"Best Score": model_tunning.best_score_

})

pd.DataFrame(scores)

In above we can see that Decision Tree is best algorithm with 92% accuracy

```
# but let&#39;s test all the algorithm on our test data set
for name, model in best_models.items():
print(name, &quot; : &quot;, model.score(X_test, y_test))
##### so we can see above when we test those trained model using test data we get random
forest is best algorithm with 96% accuracy
# let&#39;s take our final model
final_model = best_models[&quot;Random Forest&quot;]
# so here we take Random Forest model for our problem
def predict_price(year, present_price, kms, owner, diesel, petrol, individual, manual, vehicle):
vehicle_index = np.where(X.columns == vehicle.lower())[0][0]
X_pred = np.zeros_like(X.columns)
feature_list = [year, present_price, kms, owner, diesel, petrol, individual, manual]
count = 0
for f in feature_list:
X_pred[count] = f
count += 1
if vehicle_index &gt; 0:
X_pred[vehicle_index] = 1
result = final_model.predict([X_pred])
return result
predict_price(7, 9.54, 43000, 0, 1, 0, 0, 1, &quot;sx4&quot;)
# our model is working perfectly fine with around 96% accuracy
# let&#39;s save our model as binary file
import pickle
with open(&quot;model.pkl&quot;, &quot;wb&quot;) as f:
pickle.dump(final_model, f)
# let&#39;s save json file of our columns, it&#39;s use in frontend
import json
column_dict = {&quot;data_columns&quot;: X.columns.to_list()}
with open(&quot;columns.json&quot;, &quot;w&quot;) as f:
```

```
json.dump(column_dict, f)
```

7.2 FEATURE 2

```
<!DOCTYPE html>

<html lang="en">

<head>

    <meta charset="UTF-8">

    <title>Car Resale Value Prediction</title>


<link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/css/bootstrap.min.css" integrity="sha384-Gn5384xqQ1aoWXA+058RXPxPg6fy4IWvTNh0E263XmFcJlSAwiGgFAW/dAiS6JXm" crossorigin="anonymous">

<script src="https://code.jquery.com/jquery-3.2.1.slim.min.js" integrity="sha384-KJ3o2DKtIkvYIK3UENzmM7KCkRr/rE9/Qpg6aAZGJwFDMVNA/GpGFF93hXpG5KkN" crossorigin="anonymous"></script>

<script src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/1.12.9/umd/popper.min.js" integrity="sha384-ApNbgh9B+Y1QKtv3Rn7W3mgPxhU9K/ScQsAP7hUibX39j7fakFPskvXusvfa0b4Q" crossorigin="anonymous"></script>

<script src="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/js/bootstrap.min.js" integrity="sha384-JZR6Spejh4U02d8jOt6vLEHfe/JQGiRRSQQxSfFWpi1MquVdAyjUar5+76PVCmYl" crossorigin="anonymous"></script>

<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css">

<script src="https://code.jquery.com/jquery-3.5.1.js" integrity="sha256-QWo7LDvxbWT2tbbQ97B53yJnYU3WhH/C8ycbRAkjPDc=" crossorigin="anonymous"></script>

<link rel="shortcut icon" href="{{ url_for('static', filename='car.ico') }}" type="image/x-icon">

</head>

<body style = "background-image: url({{ url_for('static', filename='car_price_prediction.jpg') }});
background-repeat: no-repeat;
```

```
background-size: 1600px 980px;">


    <div class="container">
        <br><br>
        <div class="card" style="background-color:#e7e6e5">
            <div class="card-header">
                <div style="background-color:white">
                    <br>
                    <h2 class="ml-5" style="color:red"><img src="{{ url_for('static', filename='car.png') }}">     <b>Car Resale Value Prediction</b></h2>
                    <br>


                    <div class="row ml-2">
                        <div class="col-sm-5">
                            <label><b>Select Vehicle:</b></label>
                        </div>
                        <div class='col-sm-5'>
                            <select id="vehicle" onchange="vechicle_change()" name="vehicle" class="form-control " style="width:300px">
                                {% for v in vehicle %}
                                <option value="{{ v }}">{{ v.title() }}</option>
                                {% endfor %}
                                <option value="800">800</option>
                            </select>
                        </div>
                    </div>
                    <br>
                    <script type = "text/javascript">
                        function vechicle_change(){
```

```
            var vehicleObj = document.getElementById("vehicle");
            var vehicle = vehicleObj.options[vehicleObj.selectedIndex].value;

            var imageNameDisplay = document.getElementById("imageName")

            imageNameDisplay.innerHTML = "<h1>" + vehicle + "</h1>"

            var imagename = vehicle + ".jpg";

            var coverDiv = document.getElementById("imageCover")

          coverDiv.innerHTML = '<img class="card-img-top z-depth-2" src="static/vehicle images/' + imagename + '" alt="Card image cap" height="180" width="500"><br><br>'

        }
    </script>

    <div class="row ml-2">

        <div class="col-sm-4" id = "imageName">

        </div>

        <div class="col-sm-6" id = "imageCover">

        </div>
    </div>

    <div class="row ml-2">
```

```
                <div class="col-sm-5">
                    <label><b>Purchased Year: </b></label>
                </div>
                <div class="col-sm-5">
                        <input id="year" type="number" name="year" placeholder="Enter Purchased Year" class="form-control" style="width:300px">
                </div>
            </div>
            <br>

            <div class="row ml-2">
                <div class="col-sm-5">
                    <label><b>Show room price(lakh): </b></label>
                </div>
                <div class="col-sm-5">
                        <input id="show_room_price" type="number" name="show_room_price" placeholder="Enter Show Room Price" class="form-control" style="width:300px" min = "0" step="0.01">
                </div>
            </div>
            <br>

            <div class="row ml-2">
                <div class="col-sm-5">
                    <label><b>How many Kilometers Drived?: </b></label>
                </div>
                <div class="col-sm-5">
                    <input id="kilometers" type="number" name="kilometers" placeholder="Enter Kilometer" class="form-control" style="width:300px" min = "0">
                </div>
```

```
</div>
<br>

<div class="row ml-2">
  <div class="col-sm-5">
    <label><b>How much owners previously had the car?: </b></label>
  </div>

  <div class="btn-group btn-group-toggle col-sm-5" data-toggle="buttons" >
    <label class="btn btn-outline-dark">
                <input type="radio" name="owner" id="owner0" value="0" autocomplete="off"> 0
    </b></label>
    <label class="btn btn-outline-dark">
                <input type="radio" name="owner" id="owner1" value="1" autocomplete="off"> 1
    </b></label>
    <label class="btn btn-outline-dark">
                <input type="radio" name="owner" id="owner3" value="3" autocomplete="off"> 3
    </b></label>
   </div>
</div>
<br>

<div class="row ml-2">
  <div class="col-sm-5">
    <label><b>What is Fuel Type?: </b></label>
  </div>
```

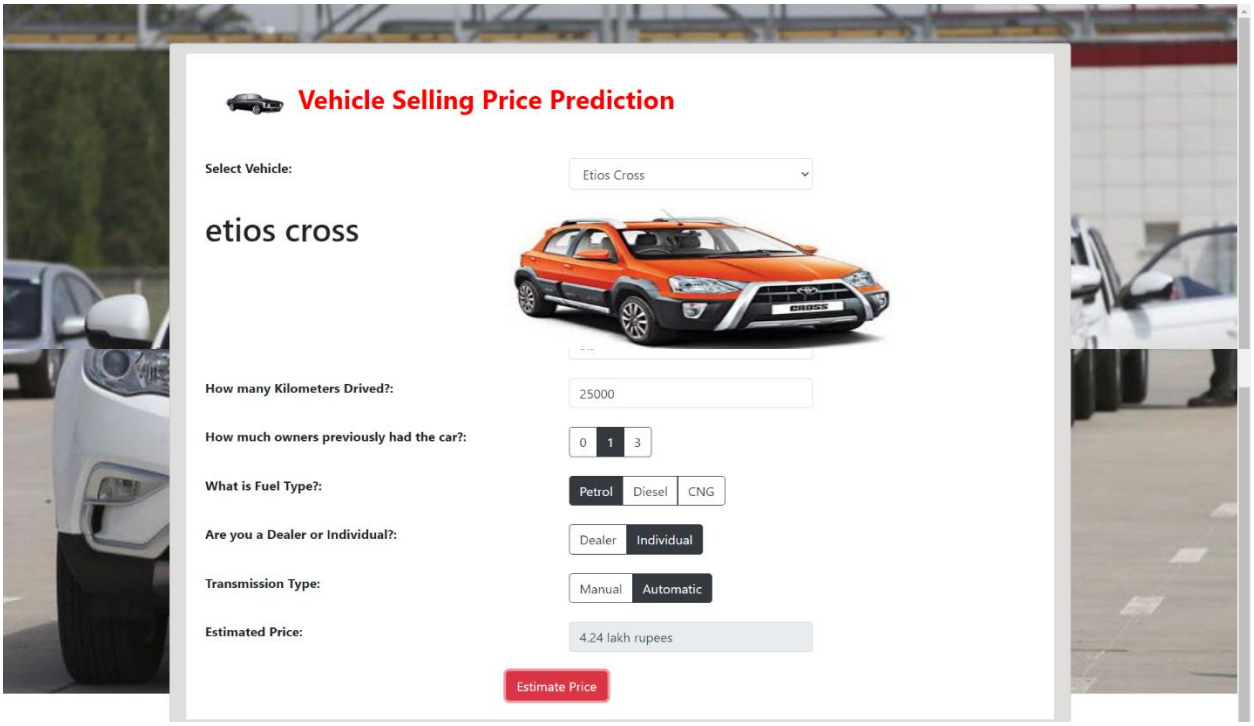
```
                <div class="btn-group btn-group-toggle col-sm-5" data-toggle="buttons" >
                  <label class="btn btn-outline-dark">
                                <input type="radio" name="fuel" id="petrol" value="petrol" autocomplete="off"> Petrol
                  </b></label>
                  <label class="btn btn-outline-dark">
                                <input type="radio" name="fuel" id="diesel" value="diesel" autocomplete="off"> Diesel
                  </b></label>
                  <label class="btn btn-outline-dark">
                 <input type="radio" name="fuel" id="CNG" value="CNG" autocomplete="off"> CNG
                  </b></label>
                 </div>
              </div>
              <br>

              <div class="row ml-2">
                 <div class="col-sm-5">
                   <label><b>Are you a Dealer or Individual?: </b></label>
                 </div>

                 <div class="btn-group btn-group-toggle col-sm-5" data-toggle="buttons" >
                   <label class="btn btn-outline-dark">
   <input type="radio" name="seller" id="dealer" value="dealer" autocomplete="off"> Dealer
  </b></label>
 <label class="btn btn-outline-dark">
<input type="radio" name="seller" id="individual" value="individual" autocomplete="off"> Individual
```

```
            </b></label>
        </div>
    </div>
    <br>
    <div class="row ml-2">
        <div class="col-sm-5">
            <label><b>Transmission Type: </b></label>
        </div>

        <div class="btn-group btn-group-toggle col-sm-5" data-toggle="buttons" >
            <label class="btn btn-outline-dark">
                <input type="radio" name="transmission" id="manual" value="manual" autocomplete="off"> Manual
            </b></label>
            <label class="btn btn-outline-dark">
                <input type="radio" name="transmission" id="automatic" value="automatic" autocomplete="off"> Automatic
            </b></label>
        </div>

    </div>
    <br>

    <div class="row ml-2">
        <div class="col-sm-10">
     <center> <button onclick="onClicked()" class = "btn btn-danger" id="btn">Estimate Price</button> </center>
            <!--
```

```
                                    <button type="submit" class="btn btn-danger"><a href="C:\Users\LENOVO\Documents\vehicle-selling-price-predictor-main\templates\b.html">Estimate Price</button>
                    -->
                    </div>
                </div>
                <br>
            </div>
            <br>
                <div class="row ml-2">
                    <div class="col-sm-5">
                        <label><b><h2>Estimated Price: </h2></b></label>
                    </div>
                    <div class="col-sm-5">
                        <input id="result" type="text" name="display_name" placeholder="Estimated Price" class="form-control" style="width:300px" disabled>
                    </div>
                </div>
                <br>
                <center>
                    <a href="https://github.com/jaysoftic" target="_blank"><i class="fa fa-github" style="font-size:24px"></i></a>    
                    <a href="https://www.linkedin.com/in/jaysoftic/" target="_blank"><i class="fa fa-linkedin-square" style="font-size:24px"></i></i></a>
                    <label class = "float-right">made by jaysoftic</label>
                </center>
            </div>
        </div>
    </div>
    <script type="text/javascript">
```

```
        function onClicked(){
            var vehicleObj = document.getElementById("vehicle");
            var vehicle = vehicleObj.options[vehicleObj.selectedIndex].value;
            var year = document.getElementById("year").value;
            var show_room_price = document.getElementById("show_room_price").value;
            var kilometers = document.getElementById("kilometers").value();
            var owner = $("input[type='radio'][name='owner']:checked").val();
            var fuel = $("input[type='radio'][name='fuel']:checked").val();
            var seller = $("input[type='radio'][name='seller']:checked").val();
            var transmission = $("input[type='radio'][name='transmission']:checked").val();
            $.post("estimatedResult", {
                year: year,
                show_room_price: show_room_price,
                kilometers: kilometers,
                owner: owner,
                fuel: fuel,
                seller: seller,
                transmission: transmission,
                vehicle: vehicle
            }, function(data, status){
                document.getElementById("result").value = data;
            })
        }
    </script>

</body>
</html>
```

OUTPUT


Vehicle Selling Price Prediction
Select Vehicle:
Etios Cross
etios cross
How many Kilometers Drived?:
25000
How much owners previously had the car?:
0
1
3
What is Fuel Type?:
Petrol
Diesel
CNG
Are you a Dealer or Individual?:
Dealer
Individual
Transmission Type:
Manual
Automatic
Estimated Price:
4.24 lakh rupees
Estimate Price

CHAPTER 8
TESTING

8.1 TEST CASES

A test case is a set of actions performed on a system to determine if it satisfies software requirements and functions correctly. Test case designing includes preconditions, case name, input conditions, and expected result. A test case is a first level action and derived from test scenarios.

• Missing values: The trained ML model requires 4 feature inputs for predicting the output. Failing which, the model throws invalid Input error. All the fields in the html form have been marked required using CSS and thus user must input all fields. Output: User must input all the fields, failing which, form shows warning message "this field needs to be filled". Thus, there can be no errors in model prediction.

• Invalid Input: The trained ML model requires only numerical input for all 4 features. Thus, if user uses symbols such as comma while input, model may throw error. To overcome the same, preprocessing script is deployed in backend which removes all unwanted characters like comma, whitespaces etc. so that model gets required input.

Output: Due to python preprocessing script, model will get the desired input and thus will give accurate prediction.

• Unseen year of purchase The model is trained with data from cars purchased since 2011 to 2020. If the user inputs details of car purchased after that i.e., 2021, model may get confused since that data is quite new and unseen to model. Output: Model has been trained with boosting algorithm and thus it gives quite accurate results with around RMSE 65,000 INR.

8.2 USER ACCEPTANCE TESTING

User Acceptance Testing (UAT), also known as beta or end-user testing, is defined as testing the software by the user to determine whether it can be accepted or not. This is the final testing performed once the functional, system and regression testing are completed. The main purpose of this testing is to validate the software against the business requirements. This validation is carried out by the end-users who are familiar with the business requirements. UAT, alpha and beta testing are different types of acceptance testing. As the user acceptance test is the last testing that is carried out before the software goes live, obviously this is the last chance for the customer to test the software and measure if it is fit for the purpose. Need for user acceptance testing arises once software has undergone Unit, Integration

and System testing because developers might have built software based on requirements document by their own understanding and further required changes during development may not be effectively communicated to them, so for testing whether the final product is accepted by client/end-user, user acceptance testing is needed.

• Developers code software based on requirements document which is their “own” understanding of the requirements and may not actually be what the client needs from the software.

• Requirements changes during the course of the project may not be communicated effectively to the developers.

CHAPTER 9

RESULTS

9.1 PERFORMANCE METRICS

Performance metrics are defined as figures and data representative of an organization’s actions, abilities, and overall quality. There are many different forms of performance metrics, including sales, profit, return on investment, customer happiness, customer reviews, personal reviews, overall quality, and reputation in a marketplace. Performance metrics can vary considerably when viewed through different industries. Performance metrics are integral to an organization's success. It's important that organizations select their chief performance metrics and focus on these areas because these metrics help guide and gauge an organization’s success. Key success factors are only useful if they are acknowledged and tracked.

CHAPTER 10

ADVANTAGES AND DISADVANTAGES

ADVANTAGES

• You can get used car at a lower price.

• You may not ha e to go into debt for used cars.

• Insurance premiums for used cars are often lower.

• Used cars give you a vintage feeling.

DISADVANTAGES

• You may miss the new-car feeling.

• Older cars area often quite loud.

• Used cars are less reliable.

• Higher maintenance cost for used cars.

• Hidden mechanical problem.

CHAPTER 11

CONCLUSION

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 pre-processing and transformation, Random Forest Regressor came out on top with 95% accuracy followed by Bagging Regressor with 88%. Each experiment was performed in real-time within the Google Colab environment. In comparison to the system's integrated Jupyter notebook and Anaconda's platform, algorithms took less training time in Google Colab. The increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on a global increase. Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of used car price prediction.

CHAPTER 12

FUTURE SCOPE

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. In the situation of increasing the prices of new cars in the market there is necessity of used car selling in every Taluka level for those people who are unable to buy high priced new cars. Therefore, there is the necessity of a car Price Prediction system which will determine the value of the car using a variety of features. The used of this model system will help to determine the accurate price of used car price prediction. With a help of most of survey paper we create a model this using linear regression algorithm and we can create a UI application for that.

CHAPTER 13

APPENDIX

SOURCE CODE: https://github.com/IBM-EPBL/IBM-Project-2831-1658483951