CAR RESALE VALUE PREDICTION

TEAM ID: PNT2022TMID13743

TEAM MEMBERS:

SWETHA S (621319205051)

SWETHA S (621319205052)

YUVARANI S (621319205062)

PAVITHRA J (621319205302)

TEAM MENTOR:

MR.R.PALANIKUMAR M.E.,

CONTENTS

CHAPTER 1

INTRODUCTION

Every day, a large number of Indian consumers sell their old cars to other buyers. The "2nd/3rd owners," etc., are the names given to these purchasers. These buyers have access to a variety of websites where they can sell their old automobiles, including cars24.com, cardekho.com, and OLX.com, but what should the cost of the car be? Machine learning techniques might be able to resolve this problem. I used a history of prior used car sales data, machine learning methods like Random Forest Regression and the powerful Python package Scikit- Learn to estimate the selling price of the used car. I also used machine learning methods like Supervised Learning. Regardless of how large or little the dataset is, the results reveal that both strategies are very accurate in prediction.

1.1 PROJECT OVERVIEW

The cost of a brand-new car is determined by the manufacturer, with some additional expenses borne by the government in the form of taxes. Customers who purchase a new vehicle can feel secure knowing that their money is being used wisely. However, as the cost of new cars rises and buyers struggle to used automobile sales are increasing globally as more people can now afford them. It is therefore vitally necessary to develop a method for calculating the value of a used car based on a number of different factors. The current method uses a tactic where a vendor chooses a price at random without the buyer knowing how much the car is worth. The seller is unsure of the car's value or what he can get for it.

To solve this issue, we have developed a model that is incredibly powerful. Regression methods are employed because they produce continuous values as opposed to put values that are categorized. As a result, it will be possible to estimate a car's actual price rather than just its price range. Additionally, a user interface that displays an automobile's pricing dependent on user input has been developed. Machine Learning, first AI refers to the study of teaching computers to function independently without being told what to do. A subset of machine learning (ML), which has emerged as the most well-known expression of the twenty-first century, is computer

based intelligence. Man-made intelligence refers to the process of making PCs appear intelligent so they can do tasks on their own. These tools are amazing. They are precise and efficient in their work. By utilising various machine learning techniques, such as supervised learning, unsupervised learning, and reinforcement learning, we develop and prepare AI techniques. The origins of machine learning may be traced back using design acknowledgment. Additionally, machine learning uses various grouping and relapse methods to deal with train models so they can learn independently.

A relatively recent quirk is the ability to spontaneously apply challenging numerical computations to large amounts of information and do it more quickly each time. While many AI techniques have been available for a while now and will continue to be, the ability to apply them to a lot of data repeatedly and quickly is a new development. Thus, the iterative component of AI becomes increasingly important as models are exposed to new data, which they adapt to using their capacity to learn from prior assumptions, and which is also responsible for delivering reliable, repeatable decisions and results. The three categories of AI strategies are next:

1) Supervised learning models must be given the input data and the ideal response before attempting to learn the rules that link the input data and the ideal outcome.

2) Unsupervised Learning: From datasets without annotations or biased examples, models are instructed to infer the hidden designs.

3) Reinforcement learning: To achieve a specific goal, the model or expert must cooperate with the distinct reality. The expert who upheld the demonstrations of the distinct world will receive compensation or rejection. The specialist will ultimately figure out how to do and explore the powerful world and achieve its objective in light of the rewards and disciplines it has received.

Most people are unsure of the differences between AI, ML, and DL, but in reality they all collaborate to develop a mindful model that can learn on its own and execute tasks without the need for human contact. The three spaces are examined in the table. When a machine completes tasks using the we refer to this "shrewd" behaviour as "Computerized reasoning," which involves

using a number of well-organized decisions to solve problems (calculations).AI isn't the same as machine learning, which is a subset of AI. It gives machines the ability to learn on their own and produce accurate predictions using the available data. Deep learning is a subset of AI. The data handling strategies discovered in the human brain, termed as Neural Networks, are typically used to liven up DL calculations.

DL's Neural Networks can be used to find a variety of them. Deep learning techniques are frequently used to train robots to carry out comparison tasks. Similar to how we use our brains to organise the final product and recognise designs. The conclusive information from the model is frequently the outcome. RNNs (Recurrent Neural Networks), CNNs (Convolutional Neural Networks), and other types of neural networks are examples of brain networks. Directed learning refers to the method of a computation obtained from a training dataset with a supervisor's assistance. It translates the contribution to the outcome and provides opportunities for information matches. Given that the first dataset already contains test information, the model predicts the outcome using values that were arbitrarily chosen from the test information. During the preparation phase, the Random Forest algorithm creates a variety of option trees using the Ensemble-Bagging method. The erratic backwoods selects the best alternate option from the majority of the results (trees). Both relapse and arrangement-regulated learning deficits can be consolidated with Arbitrary Forest. Irregular Forest is the best option for remote sensing, such as when ETM devices are used to take photographs of the earth's surface since it offers greater accuracy and requires less setup time.

Random Forest is used for Multiclass Object Detection as it performs better in harsh conditions. This method is used by a game control centre to monitor and mimic significant advancement. The Random Forest algorithm is trained to recognise body components and then benefits from that knowledge. It is then possible to identify the client's hands, feet, face, eyes, nose, and other body parts. "An irregular woodland is made up of a huge number of individual, cooperative choice trees. Every tree in the irregular timberland produces a class expectation and our model's prognosis is based on the classification that receives the most votes" the value of each element on the conjecture can also be determined using the Random Forest Algorithm. The irregular forests strategy is very simple to comprehend. Each tree in Random Forest is arbitrarily

picked from a subset of highlights. Because of the significant amount of change, there is less link between trees and more variability.

1.2 PURPOSE

In 2019, the Indian used car resale market was valued at $24.2 billion USD. There is an urgent need to close the gap between sellers and buyers due to the enormous demand for used cars and the shortage of experts who can evaluate the proper valuation. The goal of this research is to create a system that can impartially forecast a car's resale value based on little information such as the number of miles travelled and the year of purchase. The goal of a car resale value prediction system is to forecast the accurate worth of used cars, enabling customers to sell their vehicles remotely with unbiased valuation and without the need for human participation. In order to anticipate the car's resale value, the system only considers a small number of parameters due to the scarcity of data. The current system does not account for any physical damage to the car's body or engine when estimating its resale value because it is an online system.

Our newly created system is divided into two components: data collection and prediction utilising machine learning-based algorithms. To obtain information from the pages of the cars24.com website, we used web scraping libraries. The script executes and uses a URL to collect data from the HTML div mentioned in the code. The user must enter the URL. We have currently collected data by entering URLs for Swift Dzire vehicles in 5 cities. The web-based car resale value prediction is the second component. After preprocessing and cleaning the data from the previous stage, we trained an ML model that is based on the boosting technique. As a result, a user may forecast the resale value of his car with little information, without human assistance, and without personal inspection.

CHAPTER 2

LITERATURE SURVEY

TITLE: USED CAR PRICE PREDICTION.

AUTHOR: PrafulRane , Deep and Pandya, Dhawal Kotak

YEAR OF PUBLICATION: 2021

This study attempts to develop a model to forecast the reasonable costs of used automobiles based on a number of variables, such as engine size, fuel type, gearbox, road tax, and vehicle miles. In the used car market, this model can be advantageous to vendors, purchasers, and automobile manufacturers. Based on the data that users submit, it can eventually output a price prediction that is reasonably accurate. Machine learning and data science are utilised during the model development process. The dataset was obtained via scraping used car listings.

To attain the highest accuracy, a variety of regression techniques, including linear regression, polynomial regression, support vector regression, decision tree regression, and random forest regression, were used in the study. This project visualised the data to better comprehend the dataset before beginning to develop the model. To fit the regression and ensure its effectiveness, the dataset was partitioned and adjusted. R-square was calculated to assess how well each regression performed. The project's random forest regression had the highest R-square, coming in at 0.90416. The resulting model contains more elements of used cars while also being more accurate in its predictions when compared to earlier studies.

DISADVANATGE:

This study use a variety of models to forecast used automobile prices. Because there were only 380962 observations, the dataset was too tiny to draw any firm conclusions.

TITLE: CAR PRICE PREDICTION.

AUTHOR: T.B.N.L.Keerthana

YEAR OF PUBLICATION: 2021

Today, one of the foundations of the economy is thought to be the transportation sector. In wealthy countries, the automobile industry is referred to as the "Industry of Industries." Professionals in the sector claim that the UAE's automotive sector has expanded significantly. It represents its global prominence in addition to being the country with the automobile industry's quickest growth. Similar to most other nations, Dubai's residents and the ex-pat community who work there are becoming increasingly fond of cars. In the UAE, secondhand vehicles of all makes and models, including those from well-known brands, are available for purchase (Rizvi, 2019). The UAE's auto sector is expanding steadily, with a total industry volume (TIV) of 310,403 automobiles and a growth rate of 27%.Within the Gulf Cooperation Council, 1.49 million units were sold (GCC). The countries of the Gulf Cooperation Council will expand by 10% in 2021 when compared to the worldwide market (Research, 2020). The UAE market has increased by 19% thus far. As a result, in terms of growth rate, it is the greatest market in the world. Nowadays, almost everyone wants their own automobile, but many people choose to buy used cars due to issues with price or the state of the economy. Because used car prices depend on so many different features and conditions, it takes expertise to anticipate them accurately. Prices for used cars fluctuate on the market , therefore both buyers and sellers require an intelligence system to accurately anticipate the price.

DISADVANATGE:

In addition to having the highest r2 score, decision trees also had the lowest mean squared error and root mean square values, which demonstrate prediction error.

TITLE: USED CAR PRICE PREDICTION USING K-NEARESTNEIGHBOR BASED MODEL

AUTHOR: AshutoshDatt Sharma

YEAR OF PUBLICATION: 2021

Predicting the Price of Used Car Using Machine Learning Techniques is the first study. The use of supervised machine learning techniques to forecast the cost of secondhand vehicles in Mauritius is examined in this research. The forecasts are supported by historical information gathered from daily publications. The predictions were made using a variety of methodologies, including multiple linear regression analysis, k-nearest neighbours, naive bayes, and decision trees. Car Price Prediction Using Machine Learning Techniques is the second paper. For the dependable and accurate prediction, a large number of unique attributes are considered. They have employed three machine learning approaches (Artificial Neural Network, Support Vector Machine, and Random Forest) to create a model for forecasting the cost of secondhand cars in Bosnia and Herzegovina.

The third paper presents a second-hand car price evaluation model using BP neural networks. The price evaluation model based on big data analysis is put forth in this research. It makes use of widely disseminated vehicle data as well as a sizable amount of vehicle transaction data to evaluate the pricing data for each type of car using the BP neural network method that has been tuned. In order to determine the price that best fits the car, it attempts to build a model for evaluating used car prices.

DISADVANATGE:

Here, there may not be any advanced device mastery tactics to validate the version with special methods to embellish the version optimization.

TITLE: CAR POPULARITY PREDICTION

AUTHOR: P.S.L.S.Mounika,

YEAR OF PUBLICATION: 2021

It is difficult to estimate a car's resale value. It goes without saying that a variety of factors affect how much secondhand cars are worth. The age of the car, its make (and model), its origin (the country of the original manufacturer), its mileage (the number of kilometres it has driven), and its horsepower are typically the most significant ones. Fuel economy is especially crucial due to the rising cost of fuel. Unfortunately, in reality, the majority of individuals do not precisely know how much fuel their automobile uses each km. Other elements include its fuel type, interior design, braking system, acceleration, number of cylinders (measured in cubic centimetres), safety index, size, number of doors, paint colour, weight, customer reviews, prestigious awards received by the car manufacturer, physical condition, whether it is a sports car, presence of cruise control, automatic or manual transmission, whether it belonged to an individual or a company, and other elements.

As we can see, there are numerous elements that affect the pricing. Unfortunately, not all of these elements can always be known, therefore the buyer must base his decision to buy at a particular price solely on a few of them. Only a small portion of the elements stated above have been taken into account in this work. Section III gives more information. The structure of this essay is as follows. A review of relevant work is provided in the section after this. The approach is described in Section III, and the various machine learning techniques used to forecast used car prices are described, assessed, and compared in Section IV.

DISADVANATGE:

Due to frequent variations in fuel prices, the use of fuel consumption per mile has a significant impact on both the price of a car and the projection process.

TITLE: CAR PRICE PREDICTION USING MACHILNE LEARNING TECHNIQUES

AUTHOR: EnisGegic,

YEAR OF PUBLICATION: 2019

With certain additional costs imposed by the Government in the form of taxes, the manufacturer sets the prices of new cars in the market. Customers who purchase a new car can rest comfortable that their investment will be worthwhile. Used car sales, however, are rising globally as a result of the rising cost of new cars and the inability of consumers to purchase new cars due to a lack of cash. A fascinating and very required issue is predicting the cost of used autos. Customers can be greatly taken advantage of by setting inflated prices for old cars, and many fall victim to this scheme. In order to correctly assess the worthiness of the car utilising a range of factors, a used car price prediction system becomes very necessary. The majority of automobiles are purchased on a lease basis, where there is an agreement between the buyer and seller, due to the unfavourable cost of cars and the migratory lifestyle of people in industrialised countries. After the contract is over, these cars are sold again. Reselling is now an essential component of life in the modern world. Given the description of used automobiles, it is difficult to forecast what they will look like in the future. A car has several different characteristics, including its age, make, origin (the country from where it was originally manufactured), mileage (the distance it has travelled), and horsepower. Fuel economy is especially crucial due to the rising cost of fuel. Other elements include the fuel type, design, braking system, cylinder volume (in cubic centimetres), acceleration, number of doors, safety rating, size, weight, and height, paint colour, customer feedback, and significant honours earned by the automaker. The price may also be affected by other options such a music system, air conditioner, power steering, cosmic wheels, and a GPS navigator.

DISADVANATGE:

The development and launch of the product were delayed as a result of this system.

2.1 EXISTING PROBLEM

According to author Sameer Chand, projections of automobile costs were made using archived information gleaned from daily newspapers. For predicting vehicle costs, they have used the implemented AI techniques. Numerous additional calculations have also been used, including numerous straight relapse calculations, k-closest neighbour calculations, gullible based calculations, and certain choice tree computations. Each of the four calculations is examined in order to determine which is the best forecasting calculation. They have encountered a few difficulties while looking at the computations, which they have in some manner managed. This paper is more concerned with the relationship between the vendor and the customer, as stated by the authors Pattabiraman. More features, such as the previously mentioned price, mileage, make, model, trim, type, chamber, litre, entryways, voyage, sound, and cowhide, are needed in order to predict the price of four-wheelers. With the aid of a factual inquiry framework and these features, the cost of the vehicle has been projected for exploratory information analysis. The primary focus of this study, according to its authors EnisGegic et al., is on gathering various information from web entryways using web scraping techniques. Additionally, those have been put up against the help of numerous AI computations to predict the expense of the car in an easy manner. They organised the value according to the numerous types of value that have already been provided. On different datasets, fake neural networks, SVMs, and arbitrary timberland calculations were used to build classifier models. Richardson provided a different methodology in his postulation study. According to his theory, vehicle manufacturers will produce stronger vehicles. He observed how conventional and crossover cars in the scraper kept their appeal for a longer period of time by employing various relapse techniques. This utilises natural processes and moreover helps to provide huge effectiveness while using energizers. In this paper, Wu et al. demonstrate vehicle value forecast using a neurofluffy information-based architecture. They expected a model that has comparable results to the basic relapse model by taking into account the accompanying attributes, such as brand, year of production, and type of motor. They also created a specialised framework called ODAV (Optimal Distribution of Auction Vehicles) because it is common for vehicle dealers to sell used cars at the end of the rental year. This framework providesinsightsinto

the greatest prices for autos as well as the locations where all of those prices found.

2.2 REFERENCES

[1]. EnisGegic, Becker Isakovic, Dino Keco, ZerinaMasetic, and Jasmin Kevric 2019, Car Price Prediction using Machine Learning Techniques, the time interval has a significant impact on the pricing of cars in Bosnia and Herzegovina.

[2]. K.Samruddhi, & Kumar, D. R. (2020, September). Used Car Price Prediction using K-Nearest Neighbor Based Model. International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE), 4(3), 686-689.

[3] Richardson, M. S. (2019). Determinants of used vehicle resale value.

[4] Dhawal Kotak, PrafulRane, Deep Pandya(2021), Prediction of the price of a used car was developed. In the system, there are two basic phases: Traning and Testing model.

[5] Rithvik Raj Mekala, 2022, Prediction of Price forCars Using Machine Learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue VI June 2022- Available at www.ijraset.com.

[6]. Praful Rane1, Deep Pandya2, Dhawal Kotak3 “Used car price prediction “;International Research Journal of Engineering and Technology (IRJET). (2021)

[7].katamagrahari1, Ayushchobey2, Mamurkhan3, Manas srivasthava4” Car price prediction using machine learning” JUNE 2021.

[8].Abhishek pandey1, Vanshika Rastogi2, Shanika Singh “ Car selling price prediction using random forest Machine learning Algorithm, MAY 2021.

[9]."SameerchandPudaruth, “Predicting the Price of Used Cars using Machine Learning Techniques” ;(IJICT 2014).

[10].SAS Academy website - https://www.sas.com/en_in/insights/analytics/machine-learning.html.

2.3 PROBLEM STATEMENT DEFINITION

Almost all tasks that involve estimation now employ machine learning as a tool. Regression analysis is used by businesses like Cars24 and Cardeho.com to estimate used automobile prices. Therefore, we must create a model to calculate the cost of secondhand autos. The model should input parameters relevant to cars and produce a selling price. The following characteristics are what determine a used car's selling price:

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

Regression analysis can be used to resolve this supervised learning challenge. Based on the features of the given car, we must forecast the selling price of the vehicle. The selling price of a car is our target or dependent variable in a supervised regression issue, which requires labeled data. Other characteristics are all independent variables.

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

- Liner Regression
- Decision Tree Regressor
- Support Vector Regressor
- KNN Regressor

- Random Forest Regressor

Although linear models are less complicated and easier to understand, they do a poor job of limiting outliers. On nonlinear datasets, linear models do not perform well. In these scenarios, Random Forest Regressor and XGBoost Regressor nonlinear regression algorithms outperform other nonlinear regression algorithms at fitting nonlinear data. In this, the Random Forest Regressor will be used to forecast the selling price of automobiles.

CHAPTER 3

IDEATION & PROPOSED SOLUTION

The process of ideation entails coming up with new concepts and solutions through exercises like sketching, prototyping, brainstorming, writing in the head, coming up with the worst possible idea, and a variety of other ideation approaches. The third step of the Design Thinking process is also known as ideation. The ideation stage of a Design Thinking project is frequently the most exciting since it aims to produce a huge number of ideas that the team can then sort through and narrow down to the best, most useful, or most creative ones in order to inspire new and improved design solutions and products.

3.1 EMPATHY MAP CANVAS

Teams can utilise an empathy map as a collaborative tool to learn more about their clients. An empathy map can depict a group of users, such as a consumer segment, in a manner similar to user personas. The agile community has embraced the empathy may, which was first developed by Dave Gray.

To each section, everyone would put at least one sticky. You could ponder questions like:

- What would the user be feeling or thinking? What are a few of their concerns and goals?
- What are the chances that the user's boss, coworkers, and friends will say about our product when they are using it? What sounds would the user in these situations hear?
- What would the user experience like when utilising our product in their setting?
- What may a customer be doing or saying while utilising our product? In either a public or private situation, how would that differ?
- What concerns or pain points do users have when utilising our product?

• What benefits might our product's users derive from using it?

Fig.3.1.1EMPATHY MAP CANVAS

3.2 IDEATION & BRAINSTORMING

In a brainstorming session, participants get together (either physically or online) to debate, generate, articulate, and record ideas for solving issues or coming up with new things. A major firm, for instance, that just discovered it is the target of a significant lawsuit could wish to convene top executives for a brainstorming session on how to publicly address the lawsuit being filed. In a brainstorming session, participants are encouraged to freely share any ideas that may come to mind.

According to the theory, by coming up with a lot of ideas, the brainstorming group is more likely to find a workable solution to the problem they are trying to solve. With the creation of various brainstorming software tools, such Bright idea and Idea wake, the distinction between ideation and brainstorming has gotten a little bit more hazy. These software tools are made to inspire staff members to come up with fresh suggestions for enhancing business operations and, eventually, bottom-line profitability.

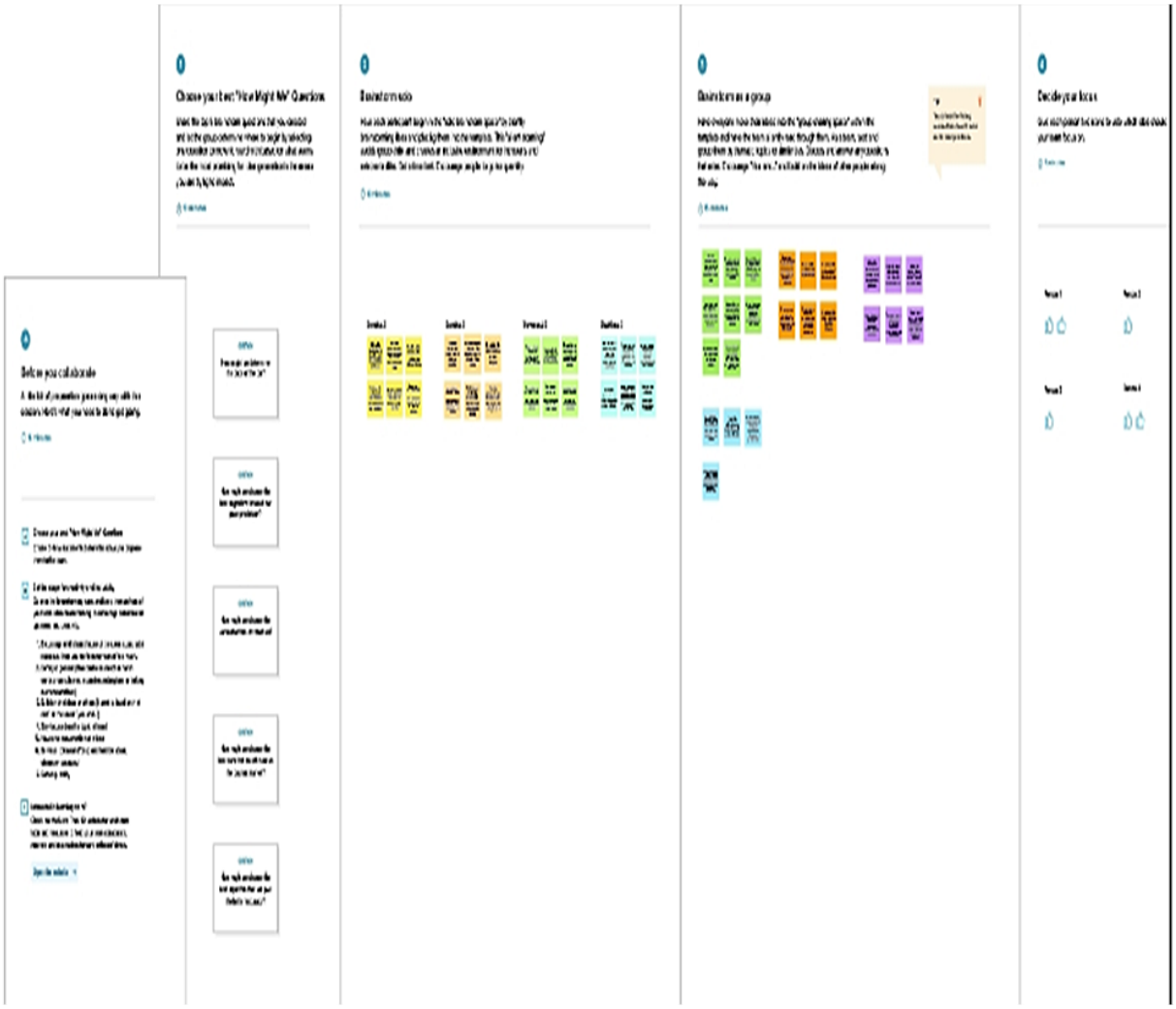


Fig.3.2.1 IDEATION & BRAINSTORMING

3.3 PROPOSED SOLUTION

1. PROBLEM STATEMENT(PROBLEM TO BE SOLVED)

To develop a method for making predictions about the cost of secondhand cars.

2. IDEA/SOLUTION DESCRIPTION

We suggested a clever, flexible, and potent system that relies on using relapse computations to predict the vehicle's resale value. Leading companies gather enormous amounts of data every day to use as a resource for business choices and solutions. The demand for data scientists and analysts is dramatically rising as a result of this enormous volume of data. Artificial intelligence and machine learning are now used in practically every industry. To remove manual interventions, businesses are implementing intelligent AI solutions in their products.

3. NOVELTY/UNIQUENESS

Used car dealers have informative advantages over statistical surveying firms, enabling them to determine resale costs with even more precision. This implies that businesses have an incentive to invest in internal decision-making systems. Instead of assembling their evaluation criteria with regard to remotely developed lingering esteem scales.

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. Competitive interest rates indicate that lower amount has to be repaid 17 towards Equated Monthly Installments (EMIs). You may choose to borrow a used car loan and buy a pre-owned car without any financial difficulties. Given that cars are now needed for everyday travel, investing in a pre-owned car is a wise decision.

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

The phrase "problem-solution fit" refers to the fact that the fundamental issue that gave rise to a business idea is genuinely present, and the suggested solution is effective in resolving it.

Fig.3.4.1 PROBLEM SOLUTION FIT

CHAPTER 4

REQUIREMENT ANALYSIS

Determining user expectations for a new or modified product is the process known as requirements analysis, sometimes known as requirements engineering. These characteristics, also known as criteria, must be precise, pertinent, and quantitative. Such specifications are sometimes referred to as functional requirements in software engineering. Project management includes requirements analysis as a key component. In order to resolve conflict or ambiguity in requirements as demanded by different users or groups of users, avoid feature creep, and document every step of the project development process from beginning to end, requirements analysis requires frequent communication with system users.

Instead of attempting to shape user expectations to match the requirements, effort should be focused on ensuring that the end system or product adheres to client needs. Requirements analysis is a collaborative endeavour that requires knowledge in hardware, software, and human factors engineering in addition to interpersonal skills. The Requirements Analysis Phase's goal is to turn the needs and high-level requirements specified in earlier phases into requirements that are clear, complete, consistent, traceable, and approved by all relevant stakeholders.

4.1 FUNCTIONAL REQUIREMENT

Functional requirements specify what a system should be able to do through calculations, technical details, data manipulation and processing, and other specific functionality. The use

cases that are used by the system to implement the functional requirements are reflected in the behavioural requirements.

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.
FR-6	Algorithm required	Using the machine learning algorithms, the final result will be predicted.

4.2 NON-FUNCTIONAL REQUIREMENT

A non-functional requirement (NFR) is a requirement that, rather than describing specific behaviours, sets criteria that can be used to assess how well a system performs. Functional requirements, on the other hand, define particular behaviours or functions. 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 demonstrates the path that information takes through a system or process. Data inputs, data outputs, data repositories, and the numerous sub-processes that the data go through are all included. Standardized symbols and terminology are used to construct DFDs, which describe numerous entities and their relationships.

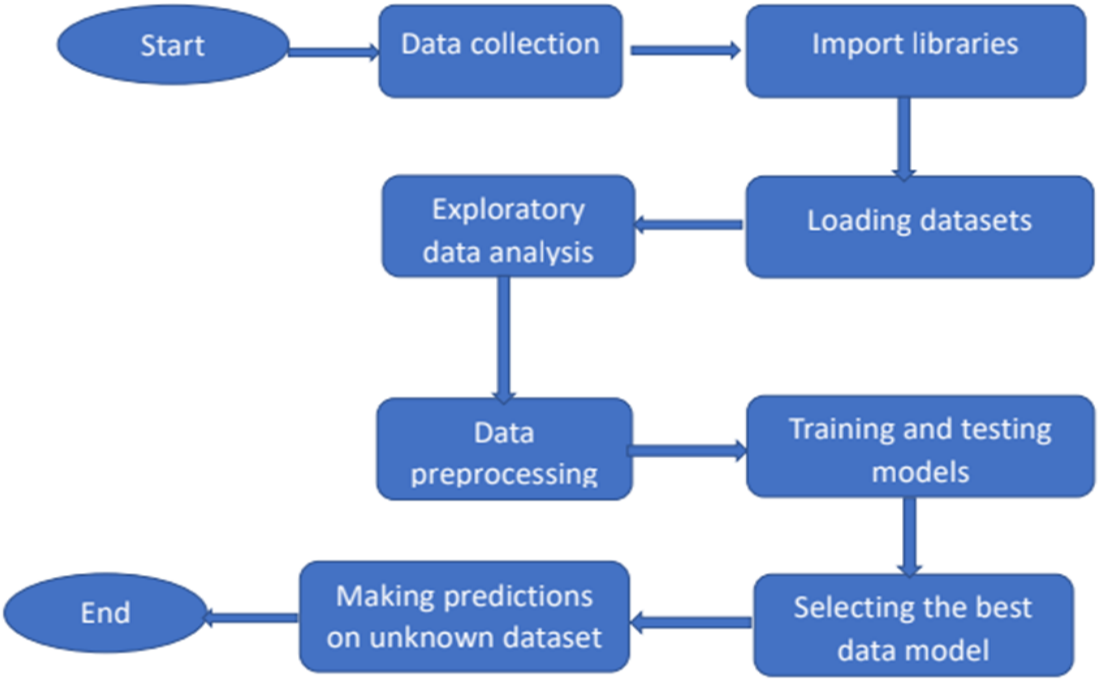


Fig.5.1.1DATA FLOW DIAGRAMS

5.2 SOLUTION & TECHNICAL ARCHITECTURE

A solution architecture diagram could actually be a collection of diagrams outlining different layers of the architecture, depending on how complicated the deployment is. The diagram makes it simple to understand how the data you gather about the environment relates to both the physical and logical decisions you make for your design.

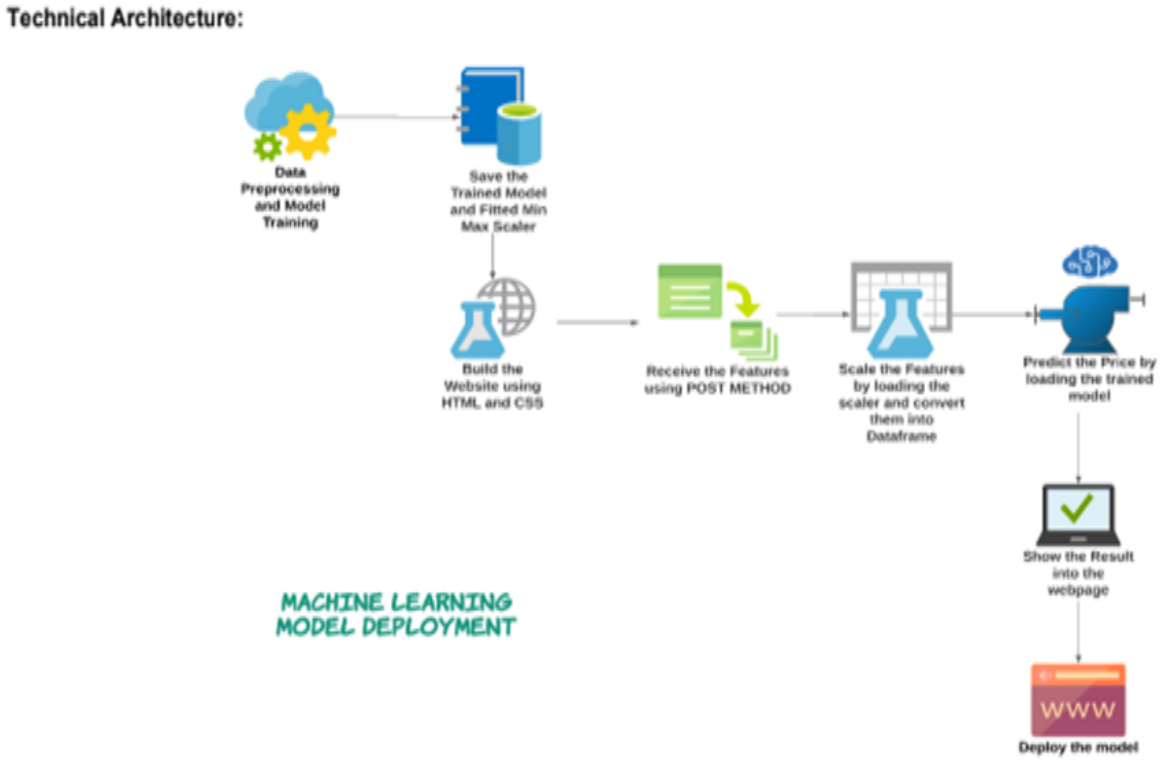


Fig.5.2.1 TECHNOLOGY STACK

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript
2.	Application Logic-1	Logic for a process in the application	Python , Linear regression
3.	Application Logic-2	Logic for a process in the application	IBM Watson STT service
4.	Application Logic-3	Logic for a process in the application	IBM Watson Assistant
5.	Database	Data Type, Configurations etc.	CSV file
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloudant etc.
7.	File Storage	File storage requirements	Local Filesystem
8.	Machine Learning Model	Purpose of Machine Learning Model	Linear regression model
9.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration :	Local, Cloud Foundry, Kubernetes, etc.

Table-2: Application Characteristics:

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

The smallest piece of work in an agile system is a user story. It is a final objective, not a feature, as seen through the eyes of a software user. A user story is a casual, all-inclusive description of a software feature written from the viewpoint of the client or end user. A user story's objective is to describe how a piece of work will provide the customer with a specific value. Keep in mind that customers don't always refer to typical exterior end users; they can also refer to internal consumers or other team members within your firm. User stories are short, straightforward statements that describe the desired result. They provide specifics. Requirements are added after the team has approved them.

User Stories

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

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (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
Administrator	User account control	USN-1	Responsible for caring out the administration process	Manage the total team	High	Sprint-1

CHAPTER 6

PROJECT PLANNING & SCHEDULING

Planning - Planning is the process of determining what supplies and resources will be needed to satisfy current and anticipated demand. This stage is essential to making sure you have the necessary supplies and resource capacity on hand to fulfill your orders on time. This element relates to the "what" and "how" of any project: precisely what must be accomplished and how it will be done.

Scheduling - Scheduling pertains to establishing establishing the time of the utilisation of particular organisational resources is referred to as scheduling. In manufacturing, scheduling entails creating schedules for personnel, machinery, and supplies. By allocating the proper resources to finish the production plan within a certain time frame, it addresses the "when" of a project.

Your facility will be able to cut expenses, boost productivity, and deliver goods on time if you create efficient production schedules. It is crucial to have a production plan that is in line with the resource and material scheduling process in order to develop precise and realistic production plans that enable manufacturers to respond quickly to changes. Any deviation or difference between the planning and scheduling processes leads to inefficiencies, which can be expensive for your company. The cost increases as the divergence increases.

6.1 SPRINT PLANNING & ESTIMATION

PLANNING:

The team decides what it will develop and how it will build it during the sprint planning phase. After breaking user stories down into tasks and performing task-level estimation, the team commits to the Sprint target. The Product Owner, Scrum Master, and Team coordinate sprint planning. Each project in Scrum is divided into sprints, which are time chunks that are typically

2-4 weeks long. The Scrum Team, Scrum Product Manager, and Scrum Master gather for a sprint planning meeting to decide which backlog items will be tackled during the following sprint.

ESTIMATION

During the Sprint Planning Meeting, the entire team estimates in Scrum projects. The goal of the estimation would be to prioritise the User Stories for the Sprint and assess the team's capacity to complete them inside the Sprint's Time Box.

The prioritised User Stories are moved to the top of the Product Backlog by the Product Owner, who also makes sure they are clear and can be estimated. The Scrum Team will take care to choose the User Stories for the Sprint based on the size of the Product Increment and the effort necessary for the same, as the Scrum Team as a whole is accountable for the delivery of the product increment. User Story Points are used to estimate the size of the product increment. Once the size n has been established, the effort—referred to as productivity—is approximated using historical data.

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
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Swetha S
	Dashboard					

Fig.6.1.1 SPRINT PLANNING & ESTIMATION

6.2 SPRINT DELIVERY SCHEDULE

Sprints are time-limited events, thus it's important to cut down on wastage during planning and production. And this is the very situation in which sprint scheduling comes into play. If you're not familiar, a sprint schedule is a written summary of the entire sprint planning process. It's one of the initial steps in the agile sprint planning process, and it calls for sufficient investigation, preparation, and coordination. When there are too many schedules created by a team, problems can arise. Conflict can result from this, and projects may get derailed in the middle of their cycles. One schedule makes sense to make sure everything proceeds as planned.

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 plots the amount of work remaining to perform against the amount of time. In agile software development approaches like Scrum, it is frequently employed. Burn down charts, however, can be used for any project that makes observable progress over time. A burn down chart typically has time along the horizontal axis and the amount of unfinished work on the vertical axis. When estimating when all of the work will be finished, it is helpful. The Development Team updates the Sprint Burn Down and plans the day's remaining tasks during the Daily Scrum.

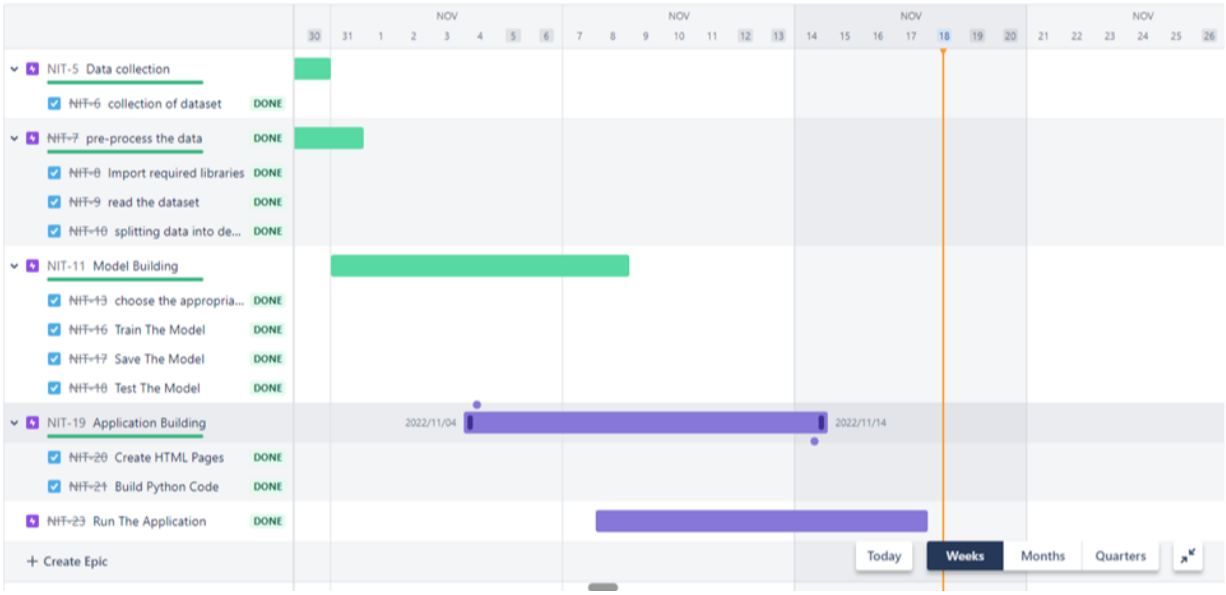


Fig.6.3.1 REPORTS FROM JIRA

CHAPTER 7

CODING AND SOLUTIONING

7.1 FEATURE 1

import pandas as pd

importnumpy as np

df = pd.read_csv("cardata.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

fromdatetime 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_TypeSeller_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't need that original text data columns so let's drop it

df.drop(columns = ["Fuel_Type", "Seller_Type", "Transmission"], inplace = True)

let's get a corelation of our data

importmatplotlib.pyplot as plt

importseaborn as sns

%matplotlib inline

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

sns.heatmap(df.corr(), annot = True)

plt.show()

let's encode Car_Name text data using one hot encoding

for this let'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't need Car_Name columns so let's drop it

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

now let's create a feature matrix X and target vector y

X = df.drop(columns="Selling_Price")

y = df.Selling_Price

let's find important feature using ExtraTreesRegressor model

fromsklearn.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

fromsklearn.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

fromsklearn.linear_model import LinearRegression

fromsklearn.tree import DecisionTreeRegressor

fromsklearn.ensemble import RandomForestRegressor

let's create a dict for paramer

algos = {

"LinearRegression": {

"model": LinearRegression(),

"param":{

"normalize":[True, False]

}

},

"DecisionTree": {

"model": DecisionTreeRegressor(),

"param": {

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

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

}

},

"RandomForest": {

"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 hyperparamtertuninig

```
fromsklearn.model_selection import RandomizedSearchCV

best_models = {}

scores = []

formodel_name, values in algos.items():

model_tunning = RandomizedSearchCV(values[&quot;model&quot;], values[&quot;param&quot;], n_iter=10, cv = 5,

n_jobs = -1)

model_tunning.fit(X_train, y_train)

best_models[model_name] = model_tunning

scores.append({

&quot;Model&quot;: model_name,

&quot;BestParameters&quot;: model_tunning.best_params_,

&quot;BestScore&quot;: 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;RandomForest&quot;]

# so here we take Random Forest model for our problem

defpredict_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
ifvehicle_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
importjson
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"><imgsrc="{{ 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">
    functionvechicle_change(){
        varvehicleObj = document.getElementById("vehicle");
        var vehicle = vehicleObj.options[vehicleObj.selectedIndex].value;

        varimageNameDisplay = document.getElementById("imageName")

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

        varimagename = vehicle + ".jpg";

        varcoverDiv = 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="btnbtn-outline-dark">
                                <input type="radio" name="owner" id="owner0" value="0" autocomplete="off"> 0
                    </b></label>
                    <label class="btnbtn-outline-dark">
                                <input type="radio" name="owner" id="owner1" value="1" autocomplete="off"> 1
                    </b></label>
                    <label class="btnbtn-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="btnbtn-outline-dark">
                    <input type="radio" name="fuel" id="petrol" value="petrol" autocomplete="off"> Petrol
        </b></label>
        <label class="btnbtn-outline-dark">
                    <input type="radio" name="fuel" id="diesel" value="diesel" autocomplete="off"> Diesel
        </b></label>
        <label class="btnbtn-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">
```

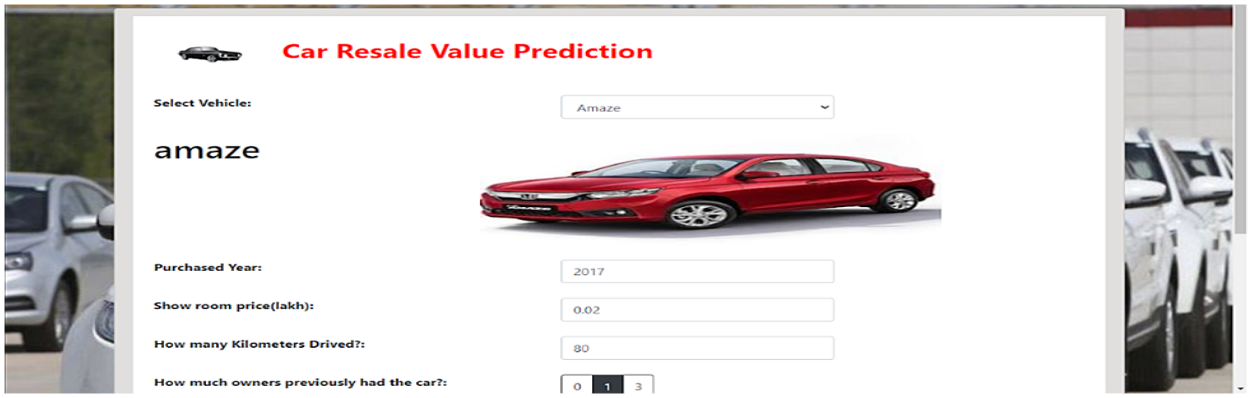
```
                    <label><b>Are you a Dealer or Individual?: </b></label>
                </div>

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

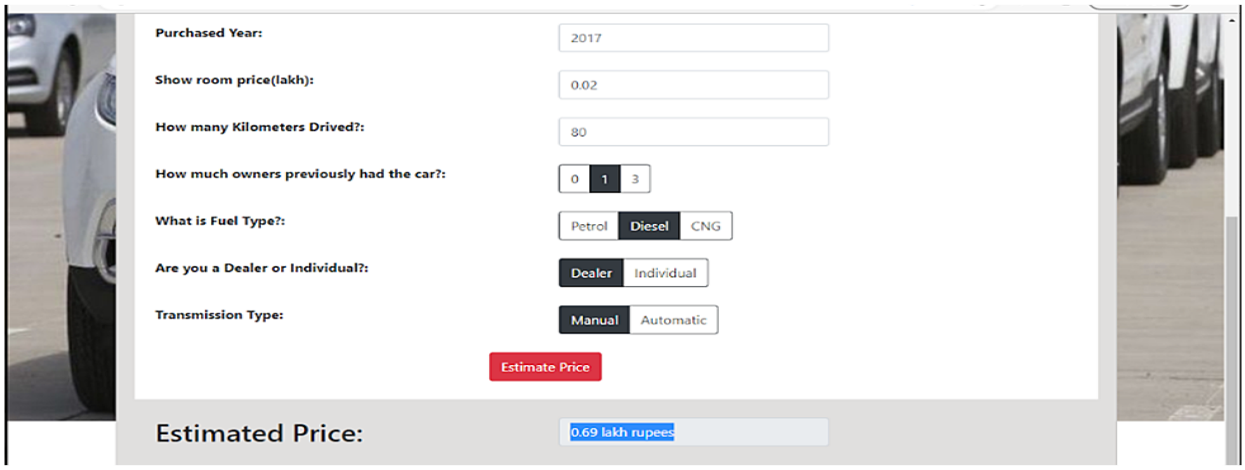
```
            </b></label>
          </div>

        </div>
        <br>

        <div class="row ml-2">
          <div class="col-sm-10">
                        <center><button onclick="onClicked()" class = "btnbtn-danger" id="btn">Estimate Price</button></center>
          <!--
                                    <button type="submit" class="btnbtn-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">
        functiononClicked(){
            varvehicleObj = document.getElementById("vehicle");
            var vehicle = vehicleObj.options[vehicleObj.selectedIndex].value;
            var year = document.getElementById("year").value;
            varshow_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


Car Resale Value Prediction
Select Vehicle:
Amaze
amaze
Purchased Year:
2017
Show room price(lakh):
0.02
How many Kilometers Drived?:
80
How much owners previously had the car?:
0
1
3


Purchased Year:
2017
Show room price(lakh):
0.02
How many Kilometers Drived?:
80
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
Estimate Price
Estimated Price:
0.69 lakh rupees

CHAPTER 8

TESTING

8.1 TEST CASES

A test case is a series of operations carried out on a system to see if it complies with software requirements and operates properly. Preconditions, case name, input requirements, and anticipated outcome are all included in test case design. An activity at the first level, test cases are derived from test scenarios.

- **Absence of values**: Four feature inputs are needed by the trained ML model to predict the output. If that doesn't work, the model returns an incorrect input error. The user must fill out every field because every one of the html form's necessary fields has been marked as such using CSS. Output: The user must fill out every field; otherwise, the form will display a warning that states, "This field has to be filled." Therefore, there can be no model prediction mistakes.

- **Invalid Input**: For all 4 characteristics, the trained ML model only needs numerical input. Thus, the model may generate an error if the user inputs symbols like commas. Preprocessing script is deployed in the backend to eliminate all undesirable characters, such as commas and whitespace, in order to ensure that the model receives the input it needs.

- **Output**: Because of the preprocessing script written in Python, the model will receive the correct input and produce accurate predictions.

- **Unseen year of purchase**: With data from vehicles acquired between 2011 and 2020, the model was trained. Since that data is quite recent and unfamiliar to the model, it may become confused if the user enters information about a car they bought after that, in 2021.

- **Output**: Because the model was trained using the boosting approach, it provides results that are fairly accurate, with an RMSE of only about 65,000 INR.

8.2 USER ACCEPTANCE TESTING

Beta testing or end-user testing, commonly referred to as user acceptance testing (UAT), is the process of having users test software to evaluate if they will accept it or not. Once the functional, system, and regression testing is finished, this is the last testing carried out. This testing's primary goal is to confirm that the software meets the necessary standards for the business. End users that are familiar with the business requirements perform this validation. Different types of acceptance testing include UAT, alpha testing, and beta testing. The user acceptance test is the last testing done before the product goes live, thus it stands to reason that this is the last opportunity for the customer to evaluate the software and determine whether it is appropriate for the task at hand. User acceptance testing is necessary after software has undergone Unit, Integration, and System testing because developers may have created the software based on the requirements document according to their own interpretation and additional required changes during development may not have been effectively communicated to them. This makes it necessary to test whether the client's or end-acceptance user's of the final product.

• Product is created by developers based on requirements documents, which may not accurately reflect the demands of the client for the software.

• The developers may not be effectively informed of requirements changes that occur during the course of the project.

CHAPTER 9

RESULTS

9.1 PERFORMANCE METRICS

Performance metrics are defined as numbers and information that are indicative of the activities, capacities, and general calibre of a company. Performance measurements can take many different forms, such as sales, profit, ROI, customer satisfaction, customer reviews, personal reviews, general quality, and reputation in the market. When evaluated through various industries, performance metrics might differ greatly. Metrics of performance are essential to the success of an organisation. Because these metrics assist direct and assess an organization's success, it is crucial that businesses choose their primary performance measures and concentrate on these areas. Important success elements are only helpful if they are recognised and monitored.

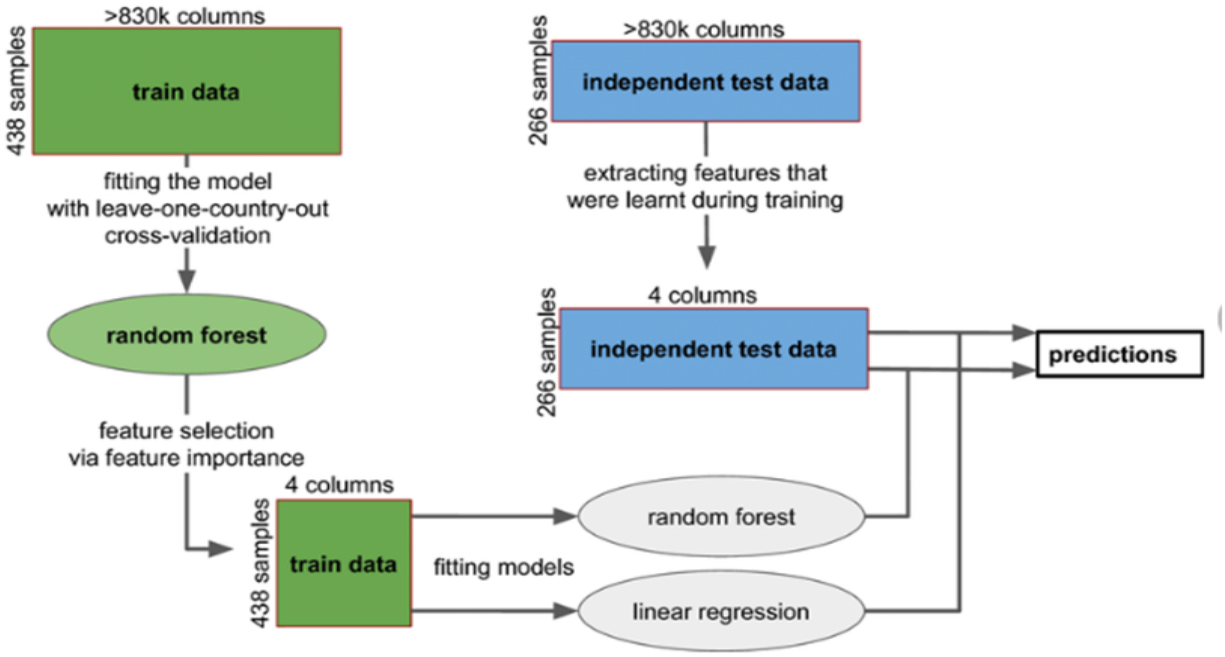


Fig.9.1.1 PERFORMANCE METRICS

CHAPTER 10

ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- You can purchase a secondhand car for less money.
- Insurance rates for used cars are frequently lower.
- You may not need to incur debt to buy one.
- Driving a used automobile gives you a retro vibe.

DISADVANTAGES

- You could miss having a brand-new car.
- Older vehicles can be rather noisy.
- Used vehicles are less trustworthy.
- Used autos have higher maintenance costs.
- Undiscovered mechanical issue.

CHAPTER 11

CONCLUSION

This research offered a scalable framework for used car price prediction in Dubai using data mining and machine learning techniques. To get the benchmark data, the Parse Hub scraping tool was used to crawl the Buyanycar.com website. Three machine learning regressors—Random Forest Regressor, Linear Regression, and Bagging Regressor are trained, tested, and evaluated in order to create an effective machine learning model. Because of the pre-processing and Bagging Regressor finished in second with 88% accuracy, and Random Forest Regressor took first with 95% accuracy after transformation. Each experiment was carried out in the Google Colab environment in real time. Algorithms in Google Colab required less training time than those in the system's integrated Jupyter notebook and the Anaconda platform. Used car sales are rising globally due to the rising costs of new cars and the customers' inability to afford to buy them. Therefore, a system that accurately assesses the value of the car utilising a range of features is urgently needed for used car price prediction. The suggested system will make it possible to estimate used automobile prices with greater accuracy.

CHAPTER 12

FUTURE SCOPE

This Initiative There will be several datasets and websites connected to the machine learning model that can supply real-time data for price prediction. Will Be Stored on GitHub or their website. Additionally, we may contribute a significant amount of data on automobile prices, which would increase the machine learning model's accuracy. As a user interface for communicating and being user-friendly with users, we are also working to develop an android app. We also intend to employ a neural network to improve the model's performance. Used car sales are necessary in every Taluka level due to the market's rising new car prices for individuals who cannot afford to purchase expensive new cars. As a result, an automobile price prediction system is required, which will assess the car's worth based on a number of criteria. The adoption of this model approach will aid in predicting used automobile prices with accuracy. We can build a UI application for this model by using most of the survey paper and the linear regression algorithm.

CHAPTER 13

APPENDIX

SOURCE CODE: http://github.com/ IBM-EPBL/IBM-Project-1563-1658397428.