**INDEX**

# 1.INTRODUCTION

## 1.1 Project Overview:

- The user interacts with the UI (User Interface) to enter the input features.
- Entered input features are analyzed by the model which is integrated.
- Once the model analyses the input, the prediction is showcased on the UI.

To accomplish this, we have to complete all the activities and tasks listed below

- Download the dataset.
- Preprocess or clean the data.
- Analyze the pre-processed data.
- Train the machine with preprocessed data using an appropriate machine learning algorithm.
- Save the model and its dependencies.
- Build a Web application using Flask that integrates with the model built.

## 1.2 Purpose:

- This can enable the customers to make decisions.
- Due to the high pricing of new cars along with the incapability of customers to invest in them, second-hand car sales are on a global increase.
- A second-hand car price prediction system is required to effectively determine the worthiness of the car using a variety of features.
- It is important to know their actual market value while both buying and selling.
- Having a fair estimate of the car's worth is a sure shot way to get the best possible value for the old car.
- As a seller, he/she wants to get the maximum price but the aim is the opposite for the buyer or the car dealer. So, to become aware of such things should be given importance.
- Need to calculate resale value of the car with the help of the correct valuation tool to know the market price or what could be the market price of the vehicle.
- To negotiate with the dealer or seller with due diligence and end up in a profitable deal.
- Estimating the best price for the car.
- Getting insight into industry rates and trends.
- Safeguarding against underhanded practices.
- Confidence for negotiations.

# 2.LITERATURE SURVEY

## 2.1 Existing Problem:

Transportation industry is one of the backbones of the economy. Almost everyone wants their own car these days, but many prefer to buy used cars or second-hand cars because of some factors like affordability and economic conditions. Used car sales are on a global increase due to the increased price of new cars and the financial incapability of the customers to buy the new

cars. The used car or second-hand market has continued to expand, as the reduction in the market of new cars. The second-hand market has created business for both buyers and sellers. Nowadays most of the people prefer to buy used cars because of the affordable price and it can also be resold after some years of usage which may get some profit. In fact the seller may not have an idea about the car's existing value in the present day's scenario or the price he should be selling the car at. Buyer too may not have an idea on the car and its value. So, the problem arises when the seller wants to fix an affordable as well as a profitable resale price for the car which would benefit both the seller and buyer. The price of used cars depends on many factors such as manufacturing year, fuel type, kilometers driven, transmission type, engine, etc., Accurately predicting the used car prices requires expert knowledge due to their nature of dependence on a variety of factors and features. Therefore, an efficient application or website built using an effective evaluation model to predict the resale value of the car is required.

**2.2 References**

[1] Sameerchand Pudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques", http://ripublication.com/irph/ijict_spl/ijictv4n7spl_17.pdf, 2014.

[2] Ashish Chandak, Prajwal Ganorkar, Shyam Sharma, Ayushi Bagmar, Soumya Tiwari, "Car Price Prediction using Machine Learning",https://www.ijcseonline.org/pub_paper/74-IJCSE-07051.pdf, 2019.

[3] K. Samruddhi, Dr. R. Ashok Kumar, "Used Car Price Prediction using K-Nearest Neighbor Based Model", https://www.ijirase.com/assets/paper/issue_1/volume_4/V4-Issue-3-686-689.pdf, 2020.

[4] S.E.Viswapriya, Durbaka Sai Sandeep Sharma, Gandavarapu Sathya Kiran, "Vehicle Price Prediction using SVM Techniques", https://www.ijitee.org/wp-content/uploads/papers/v9i8/G5915059720.pdf, 2020.

[5] Praful Rane, Deep Pandya, Dhawal Kotak, "Used Car Price Prediction", https://www.irjet.net/archives/V8/i4/IRJET-V8I4278.pdf, 2021.

[6] Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, "Car Price Prediction using Machine Learning Techniques", https://www.temjournal.com/content/81/TEMJournalFebruary2019_113_118.pdf, 2019.

[7] Abdulla Alshared, "Used Cars Price Prediction and Valuation using Data Mining Techniques", https://scholarworks.rit.edu/cgi/viewcontent.cgi?article=12220&context=theses, 2021.

[8] Ashutosh Datt sharma, Vibhor sharma, Sahil Mittal, Gautam Jain, Sudha Narang, "Predictive Analysis of Used Car Prices using Machine Learning", https://www.irjmets.com/uploadedfiles/paper/volume3/issue_6_june_2021/12071/1628083486.pdf, 2021.

[9] Aditya Nikhade, Rohan Borde, "Used Car Price Prediction and Life Span", https://iarjset.com/wp-content/uploads/2022/01/IARJSET.2021.81249.pdf, 2021.

[10] Ram Prashath R, Nithish C N, Ajith Kumar J, “Price Prediction of Used Cars using Machine Learning”,https://www.ijraset.com/research-paper/price-prediction-of-used-cars-using-machine-learning, 2022.

### 2.3 Problem Statement Definition:

With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. In many developed countries, it is common to lease a car rather than buying it outright. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, it is of commercial interest to sellers/financers to be able to predict the salvage value (residual value) of cars with accuracy.

Therefore we need an intelligent solution to predict the accurate resale value of the car and present it to the users in a web application.

## 3. IDEATION & PROPOSED SOLUTION

### 3.1 Empathy Map Canvas

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

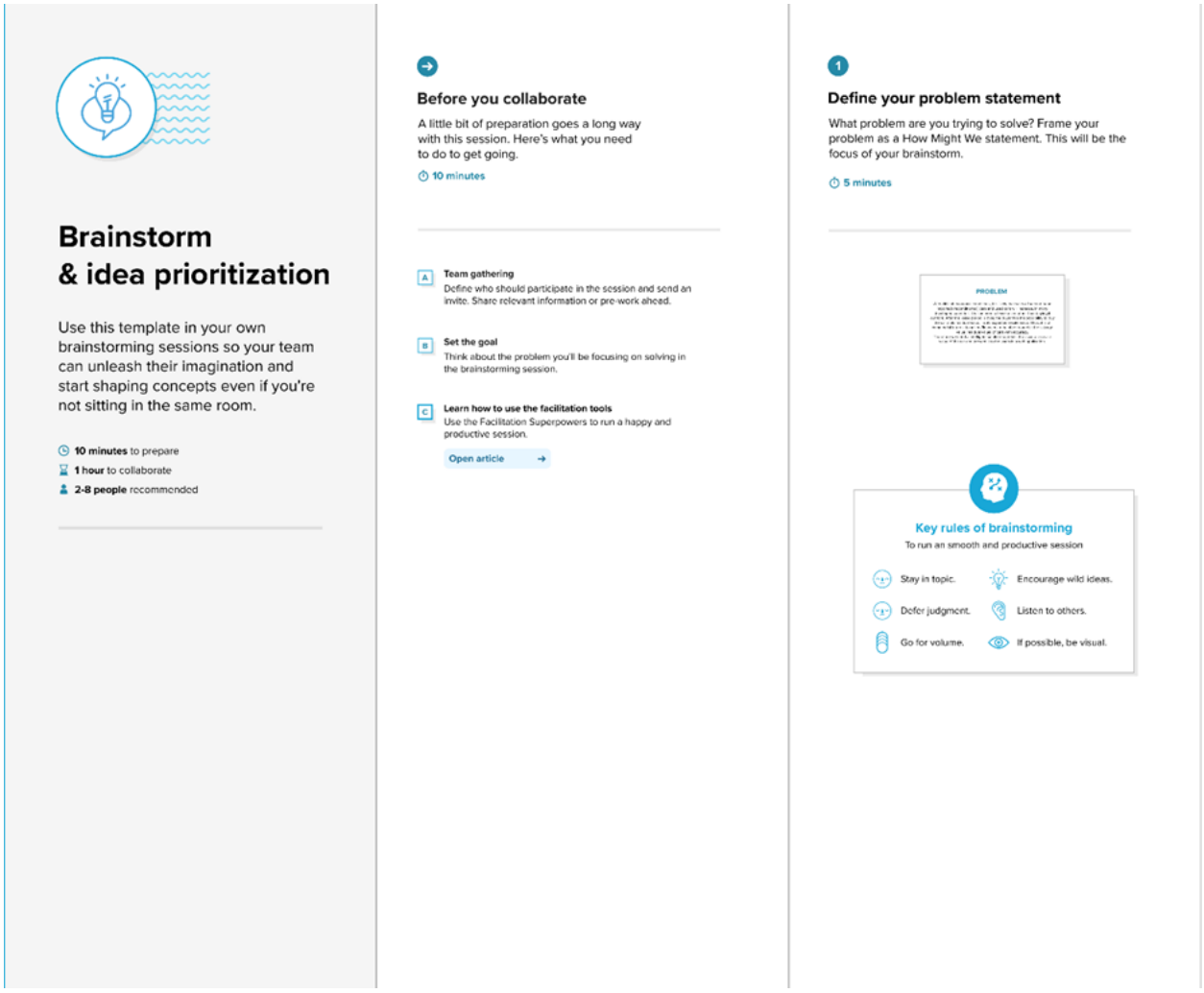
| | |
|---|---|
| **Think & Feel**<br>● Give better results?<br>● Works fast?<br>● Can I have a quick prediction?<br>● What are the benefits? | **Hear**<br>● Great idea<br>● Resale is good<br>● Cost efficiency |
| **See**<br>● Updated values<br>● Cost<br>● Better results<br>● Quick prediction | **Say & Do**<br>● Add on features<br>● Looking for better results<br>● More easy to predict<br>● Clear ideas |

| Pain | Gain |
| --- | --- |
| • Hard to find prices<br>• Sometimes late pred | • Accuracy<br>• Comfortable |

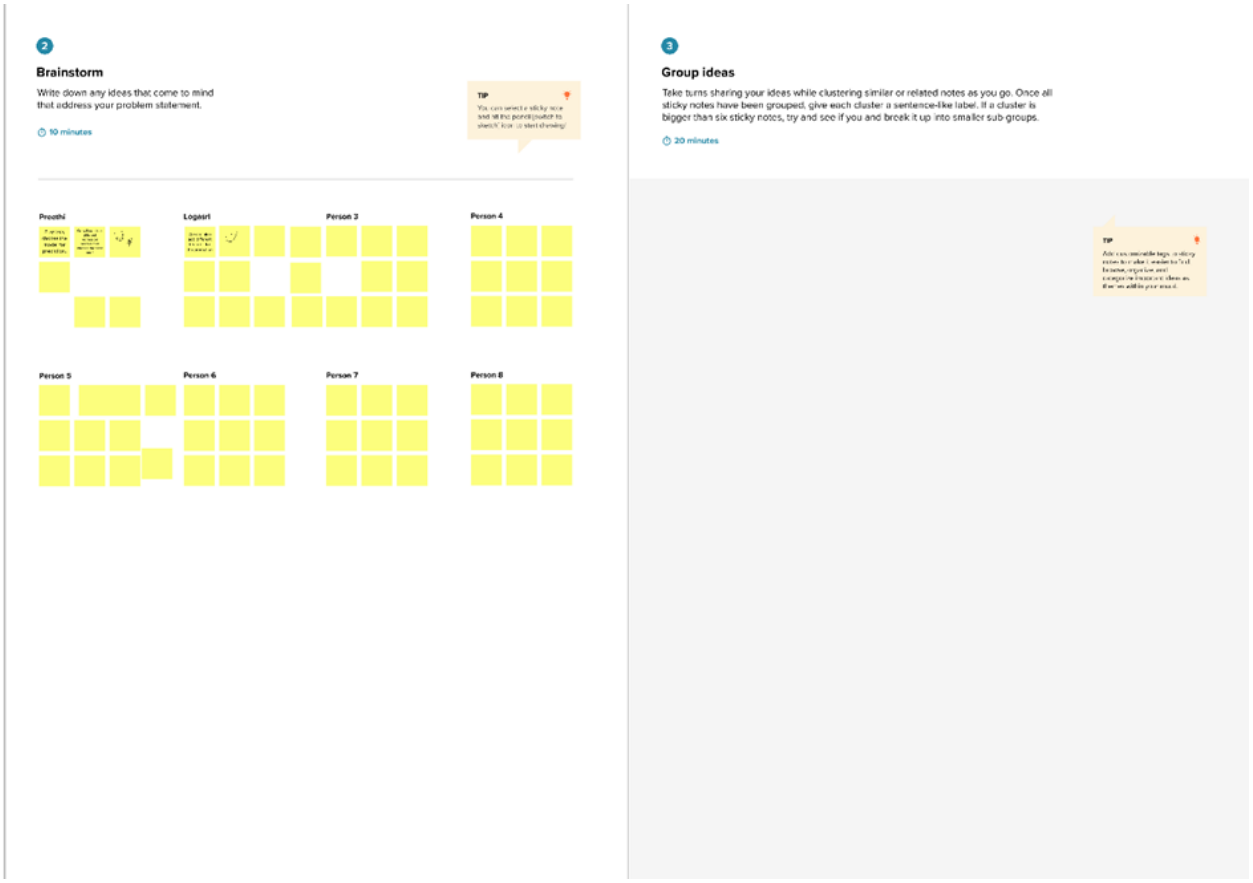
### 3.2 Ideation & Brainstorming

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

**Step-1: Team Gathering, Collaboration and Select the Problem Statement**



## Step-2: Brainstorm, Idea Listing and Grouping



## Step-3: Idea Prioritization

4

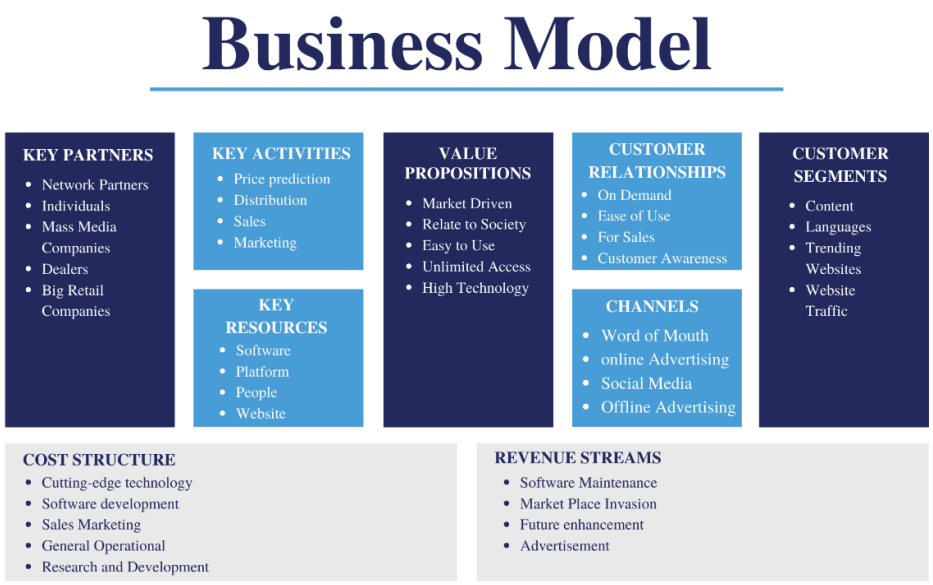
### Prioritize

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

20 minutes

### 3.3 Proposed Solution

| S.No | Parameter | Description |
|---|---|---|
| 1. | Problem Statement (Problem to be solved) | **Car Resale Value Prediction**<br>With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. In many developed countries, it is common to lease a car rather than buying it outright. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e., its expected resale value. Thus, it is of commercial interest to sellers or financers to be able to predict the salvage value (residual value) of cars with accuracy. Therefore, the problem statement is to predict the resale value of Cars. |
| 2. | Idea / Solution description | ● In order to predict the resale value of the car, we have proposed an intelligent, flexible, and effective solution which works based on a machine learning algorithm.<br>● Considering the main factors which would affect the resale value of a car, the machine learning model has been trained to give the nearest resale value of the car.<br>● We have tried various machine learning algorithms and the algorithm with the best accuracy has been taken as a solution, then it has been integrated to the website developed by us, where the user is displayed with the resale price of the car.<br>● The machine learning model that we have used is Random Forest Regression. |
| 3. | Novelty / Uniqueness | ● Most of the car resale prediction websites or applications take into account only some basic features of the car like kilometers driven, fuel type, the year when the car was bought and the showroom price of the car.<br>● But we have used some more specific parameters like number of owners who previously had the car, the car transmission type (i.e., Manual car or Automatic car) and whether the person who is willing to sell the car is an individual or dealer.<br>● The car resale price which is predicted by our model is acceptable as it gives the nearest price as output. |
| 4. | Social Impact / Customer Satisfaction | ● Customers become aware of the car value before they buy the used cars so that fraudulent activities can be prevented.<br>● Gives an idea for the sellers regarding the car resale price. |

| | | |
|---|---|---|
| | | • Our Car resale prediction website sets a price at a level that prediction attracts customers and also maximizes sales.<br>• Knowing the resale price beforehand helps the customer as well as the seller in many ways by minimizing the losses. |
| 5. | Business Model (Revenue Model) |  |
| 6. | Scalability of the Solution | • This prediction saves the time of both the customer and the seller.<br>• Does not require any installation process to use the website.<br>• The website is user friendly.<br>• Boosts the sale of used cars.<br>• Increases competitiveness of market players, by making this resale business more attractive. |

### 3.4 Problem Solution fit

The sales of second-hand imported cars and used cars is increasing nowadays. Predicting the price of used cars is an important and interesting problem. Predicting the resale value of a car is not an easy task. It is trite knowledge that the value of used cars depends on a number of factors. The value of a car drops right from the moment it is bought and the depreciation continues with each passing year. In fact, in the first year itself, the value of a car decreases by 20 percent of its initial value. The make and model of a car, total kilometers driven, overall condition of the car and various other factors further affect the car's resale value. So, it is necessary to build a model and design an application or website to estimate the price of used cars. The model should take car related parameters and output a selling price of the car. 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, etc.,

| | | |
|---|---|---|
| **1. CUSTOMER SEGMENT(S)**<br>● Consumers who are actively looking to buy an automobile.<br>● This audience segment also has people who actively look to make a purchase by frequently visiting automotive-related locations such as dealerships and visiting review and price comparing websites.<br>● When most people think of the target market for used cars, they typically think of young people who are just starting out in life and who are looking for a cheaper alternative to a new car. | **6. CUSTOMER CONSTRAINTS**<br>● Check the Car's Condition<br>● Maintenance Records<br>● Car Mileage<br>● Check Registration Certificate<br>● Car Insurance<br>● Fuel Type<br>● Transmission Type (i.e., Manual or Automatic) | **5. AVAILABLE SOLUTIONS**<br>● Prediction of resale probability.<br>● Prediction of resale value with only few parameters.<br>● Websites that predict whether the used car or second-hand car is eligible for sales or not. |
| **2. JOBS-TO-BE-DONE /PROBLEMS**<br>● Accuracy of the predicted value.<br>● Providing the users with the most accurate resale value as output.<br>● Overfitting in model.<br>● Errors in the training or prediction process.<br>● Customer Satisfaction.<br>● User Interface problems. | **9. PROBLEM ROOT CAUSE**<br>● Training the model to provide accurate results.<br>● Constantly changing trends of cars.<br>● Parameters or factors that affect the resale value of the cars. | **7. BEHAVIOUR**<br>● If the prediction is not accurate then the user tends to go in search of other websites.<br>● The customer gets irritated or stressed if the user interface design is not good and faces difficulties in working with the website.<br>● These are some of the behaviors of users that may occur generally. |

<table>
<tr>
<td>3. TRIGGERS<ul><li>Reading about innovative ideas.</li><li>Advertisements and social media promotions.</li><li>More efficient solution.</li><li>Recommendation from their well-wishers.</li></ul></td>
<td rowspan="2">10. YOUR SOLUTION<ul><li>In order to predict the resale value of the car, an intelligent, flexible, and effective solution which works based on a machine learning algorithm has been found.</li><li>Considering the main factors which would affect the resale value of a car, the machine learning model has been trained to give the nearest resale value of the car.</li><li>The machine learning model that has been used is Random Forest Regression. This gives the reasonable resale value of the car.</li></ul></td>
<td>8. CHANNELS OF BEHAVIOUR<br>8.1 ONLINE<ul><li>Reviews</li><li>Comments</li><li>Blogs</li><li>Posts</li><li>Status</li><li>Sharing information through social media sites</li></ul></td>
</tr>
<tr>
<td>4. EMOTIONS: BEFORE/AFTER<br>Before the Problem is solved:<ul><li>Stress</li><li>Frustration</li><li>Anger</li><li>Blocking</li><li>Bad opinion on the website</li></ul>After the Problem is solved:<ul><li>Feeling smart</li><li>Good opinion on the website</li><li>Trust</li><li>Boost</li><li>Happy</li></ul></td>
<td>8.2 OFFLINE<ul><li>Mouth of word</li><li>Communication</li><li>Meetings</li><li>Letters</li></ul></td>
</tr>
</table>

## 4. REQUIREMENTS

### 4.1 Functional Requirements:

**Functional requirements** are product features or functions that developers must implement to enable users to accomplish their tasks. So, it's important to make them clear both for the development team and the stakeholders. Generally, functional requirements describe system behavior under specific conditions.

A typical functional requirement will contain a unique name and number, a brief summary, and a rationale. This information is used to help the reader understand why the requirement is needed, and to track the requirement through the development of the system. The crux of the requirement is the description of the required behavior, which must be clear and readable. The described behavior may come from organizational or business rules, or it may be discovered through elicitation sessions with users, stakeholders, and other experts within the

organization.Many requirements may be uncovered during the use case development. When this happens, the requirements analyst may create a placeholder requirement with a name and summary, and research the details later, to be filled in when they are better known.

Following are the functional requirements of the proposed solution.

| FR No. | Functional Requirement (Epic) | Sub Requirement (Story / Sub-Task) |
|---|---|---|
| FR-1 | User Input Entry | Enter the year of purchase, showroom price<br>Enter the number of kilometres driven<br>Enter the previous car owners, fuel type<br>Enter the transmission type of the car |
| FR-2 | Prediction Result | Prediction results to be shown in the UI based on the input data. |
| FR-3 | Save the result | The results of the prediction to be saved in the form of pdf. |
| FR-4 | Application Accessibility | Access the application from any device, any web browser and from anywhere |
| FR-5 | User Interface | View all the elements of the application clearly without any disturbance. |
| FR-6 | Maintenance | Contact the customer care if there are any issues. |

**4.2 Non Functional Requirements:**

Non-Functional Requirements are the constraints or the requirements imposed on the system. They specify the quality attribute of the software. Non-Functional Requirements deal with issues like scalability, maintainability, performance, portability, security, reliability, and many more. Non-Functional Requirements address vital issues of quality for software systems. If NFRs not addressed properly, the results can include:

- Users, clients, and developers are unsatisfied.
- Inconsistent software.
- Time and cost overrun to fix the software which was prepared without keeping NFRs in mind.

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

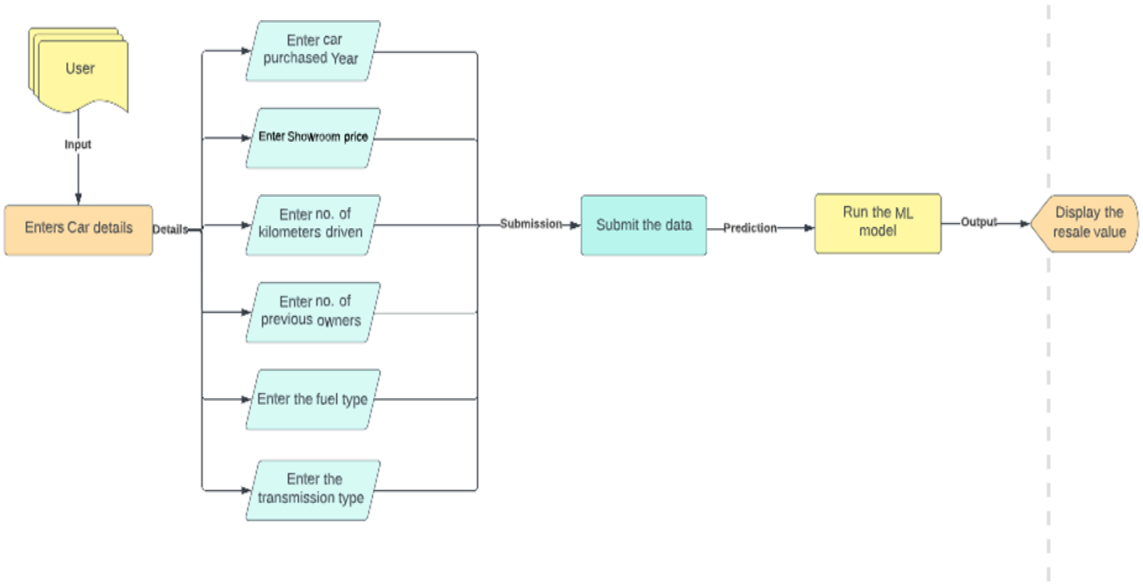
| FR No. | Non-Functional Requirement | Description |
|---|---|---|
| NFR-1 | Usability | The prediction and the display of the output must be done within 2 seconds |
| NFR-2 | Security | The software must remain resilient in the face of attacks. The software must be available and behave reliably even under DOS attacks. |
| NFR-3 | Reliability | Public can access the website 99% of the time without failure. |
| NFR-4 | Performance | The website's load time should not be more than one second for users. |
| NFR-5 | Recoverability | If a major incident happens on the website, the business must take measures to go back to being fully operational within three days. |
| NFR-6 | Security | No bots must access the website. |

## 5. PROJECT DESIGN

### 5.1 Data Flow Diagrams

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deeper into how the data is handled. They can be used to analyze an existing system or model a new one. Like all the best diagrams and charts, a DFD can often visually “say” things that would be hard to explain in words, and they work for both technical and nontechnical audiences, from developer to CEO. That’s why DFDs remain so popular after all these years. While they work well for data flow software and systems, they are less applicable nowadays to visualizing interactive, real-time or database-oriented software or systems.

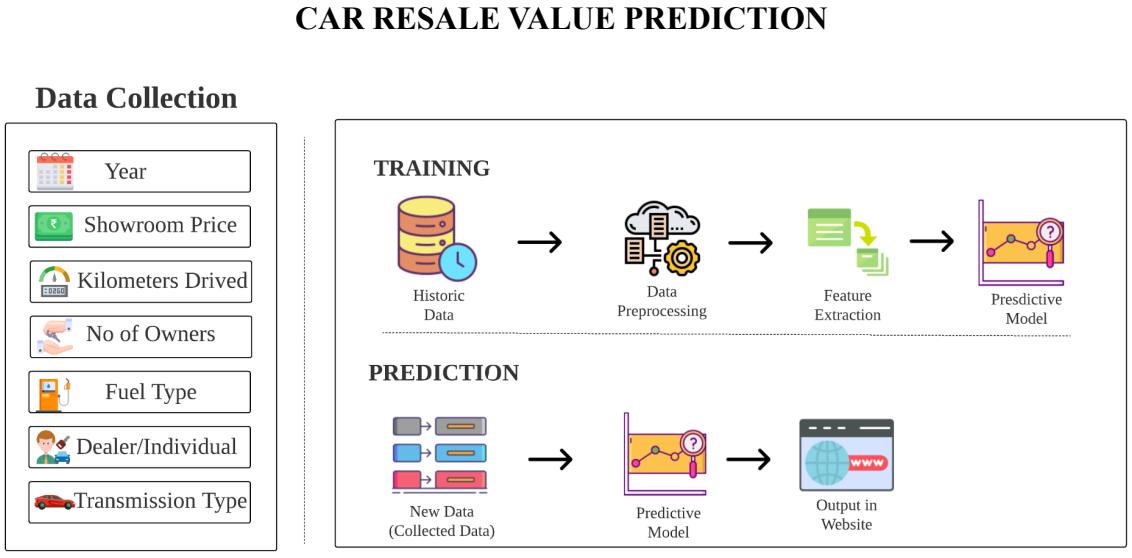
Data flow diagram for the project:




## 5.2 Solution and Technical Architecture:

### Solution Architecture:

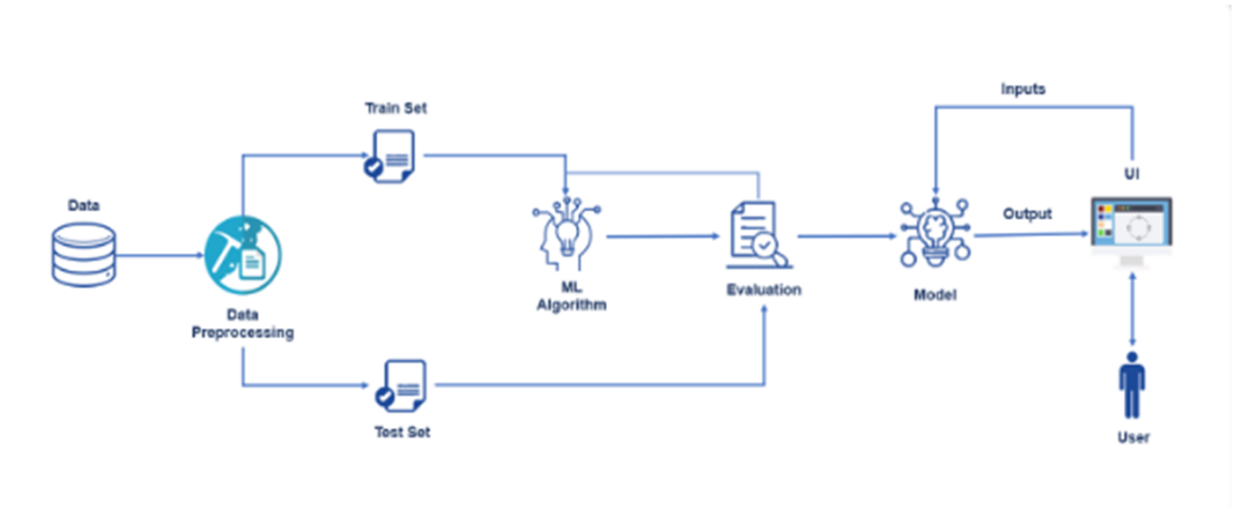
Solution Architecture is an architectural description of a specific solution. It is the practice of designing, describing, and managing solution engineering to match its specific business problems. It comprises subprocesses that draw guidance from various enterprise architecture viewpoints. Solution Architecture is the initial step taken when an organization aims to create a set of enterprise solutions, applications and processes that integrate with each other in order to address specific needs and requirements and that often lead to software architecture and technical architecture work.



**Technical Architecture:**

Technical architecture—which is also often referred to as application architecture, IT architecture, business architecture, etc.—refers to creating a structured software solution that will meet the business needs and expectations while providing a strong technical plan for the growth of the software application through its lifetime. IT architecture is equally important to the business team and the information technology team.

Technical architecture includes the major components of the system, their relationships, and the contracts that define the interactions between the components. The goal of technical architects is to achieve all the business needs with an application that is optimized for both performance and security. The technical architecture of the proposed project is as follows:




**5.3 User Stories**

A user story is an informal, general explanation of a software feature written from the perspective of the end user. Its purpose is to articulate how a software feature will provide value to the customer. 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. The user story for the project is as follows:

| User Type | Functional Requirement (Epic) | User Story Number | User Story / Task | Acceptance criteria |
|---|---|---|---|---|
| Customer (Web user) | Details to be entered | USN-1 | As a user, I can enter the year when i have purchased my car | I can enter the value in a text box. |
| | | USN-2 | As a user, I can enter the showroom price, number of kilometres driven, number of previous car owners | I can enter the value in a text box |
| | | USN-3 | As a user, I can enter the fuel type, dealer, transmission type. | I can do by selecting the options in dropdown |
| | | USN-4 | As a user, I can view the resale of the car | I can do it by clicking the submit button. |
| | | USN-5 | As a user, I must be able to save the details for future reference | I can save the details entered and the value in a pdf |
| | UI/UX | USN-1 | As a user, I can view all the elements without any visual disturbance | I can view all the required elements clearly |
| Customer (Mobile user) | UI | USN-1 | As a user, I must be able to view all the entering text boxes clearly. | I can do it if the compatibility is given. |
| | | USN-2 | As a user, I can view the output clearly | I can do it if the font size is clear |
| Customer Care Executive | Maintenance | USN-1 | As a user, I must be able to contact the maintenance team whenever any issues arise. | I can do it by sending email to the team |
| Administrator | Security | USN-1 | The customer must not be able to view how the prediction is done | The ML model must not be visible to the user. |

## 6. PROJECT PLANNING & SCHEDULING

### 6.1 Sprint Planning and Estimation

Sprint planning is an event in scrum that kicks off the sprint. The purpose of sprint planning is to define what can be delivered in the sprint and how that work will be achieved. Sprint planning is done in collaboration with the whole scrum team. In scrum, the sprint is a set period of time where all the work is done. However, before you can leap into action you have to set up the sprint. You need to decide on how long the time box is going to be, the sprint goal, and where you're going to start. The sprint planning session kicks off the sprint by setting the agenda and focus. If done correctly, it also creates an environment where the team is motivated, challenged, and can be successful. Bad sprint plans can derail the team by setting unrealistic expectations.
The following is the sprint planning and estimation for the project

| Release | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority |
|---|---|---|---|---|---|
| Sprint-1 | Details to be entered | USN-1 | As a user, I can enter the year when i have purchased my car | 2 | High |
| Sprint-1 | | USN-2 | As a user, I can enter the showroom price, number of kilometers driven, number of previous car owners | 5 | High |
| Sprint-3 | | USN-3 | As a user, I can enter the fuel type, dealer, transmission type. | 3 | High |
| Sprint-3 | | USN-4 | As a user, I can view the resale of the car | 5 | High |
| Sprint-3 | | USN-5 | As a user, I must be able to save the details for future reference | 1 | Medium |

| | | | | | |
|---|---|---|---|---|---|
| Sprint-2 | UI/UX | USN-1 | As a user, I can view all the elements without any visual disturbance | 4 | High |
| Sprint-1 | UI | USN-1 | As a user, I must be able to view all the entering text boxes clearly. | 3 | High |
| Sprint-3 | | USN-2 | As a user, I can view the output clearly | 3 | High |
| Sprint-2 | Maintenance | USN-1 | As a user, I must be able to contact the maintenance team whenever any issues arise. | 3 | Medium |
| Sprint-2 | Security | USN-1 | The customer must not be able to view how the prediction is done | 5 | High |

**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, planning, and communication.

Sprint Delivery Schedule for the proposed solution:

| **Sprint** | **Total Story Points** | **Duration** | **Sprint Start Date** | **Sprint End Date (Planned)** |
|---|---|---|---|---|
| Sprint-1 | 20 | 6 Days | 24 Oct 2022 | 29 Oct 2022 |
| Sprint-2 | 20 | 6 Days | 31 Oct 2022 | 05 Nov 2022 |

| | | | | |
|---|---|---|---|---|
| Sprint-3 | 20 | 6 Days | 07 Nov 2022 | 12 Nov 2022 |

## 7. CODING & SOLUTIONING

### 7.1 FEATURE 1

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 problems require labeled data where our target or dependent variable is the selling price of a car. All other features are independent variables.
Linear Models are relatively less complex and explainable, but linear models perform poorly on data containing the outliers. Linear models fail to perform well on non-linear datasets. In such cases, non-linear regression algorithms Random Forest Regressor perform better in fitting the nonlinear data.
This model was hence chosen to account for the large number of features in the dataset and compare a bagging technique with the following gradient boosting methods.

### 7.2 FEATURE 2

Given the evaluation parameters the Random Forest Regressor outperformed as it has the highest accuracy as well as the lowest error in all three valuation parameters.
As a result of preprocessing and transformation, Random Forest Regressor came out on with 90% accuracy.

## 8. TESTING

### 8.1 TEST CASES

- **Missing values**

The trained ML model requires few 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 the user must input all fields.

- **Invalid Input**

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

## 8.2 USER ACCEPTANCE TESTING

Acceptance testing focuses even more on the overall system features and functionality that are visible to the customer. Acceptance testing is often performed by customers to ensure customer usability and satisfaction. The purpose of this is to briefly explain the test coverage and open issues of the [ProductName] project at the time of the release to User Acceptance Testing (UAT).

1. **Defect Analysis**This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

| Resolution | Severity 1 | Severity 2 | Severity 3 | Severity 4 | Subtotal |
|---|---|---|---|---|---|
| By Design | 10 | 4 | 2 | 3 | 19 |
| Duplicate | 1 | 0 | 1 | 0 | 2 |
| External | 2 | 1 | 0 | 2 | 5 |
| Fixed | 13 | 2 | 2 | 20 | 37 |
| Not Reproduced | 0 | 0 | 1 | 1 | 2 |
| Skipped | 0 | 0 | 1 | 1 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| Won't Fix | 0 | 2 | 1 | 1 | 4 |
| Totals | 26 | 9 | 8 | 28 | 71 |

**2. Test Case Analysis**

This report shows the number of test cases that have passed, failed, and untested

| Section | Total Cases | Not Tested | Fail | Pass |
|---|---|---|---|---|
| Print Engine | 5 | 0 | 0 | 5 |
| Client Application | 13 | 0 | 0 | 13 |
| Security | 2 | 0 | 0 | 2 |
| Outsource Shipping | 4 | 0 | 0 | 4 |
| Exception Reporting | 6 | 0 | 0 | 6 |
| Final Report Output | 4 | 0 | 0 | 4 |
| Version Control | 2 | 0 | 0 | 2 |

## 9. RESULTS

### 9.1 PERFORMANCE METRICS

Performance metrics are used to track progress. Metrics give some sort of concrete answer which easily can be followed up. There are different types of metrics used for testing.

The regression model can be evaluated on following parameters:

**Regression Model:**

1. Mean Square Error (MSE): MSE is the single value that provides information about goodness of regression line. Smaller the MSE value, better the fit because smaller value implies smaller magnitude of errors.

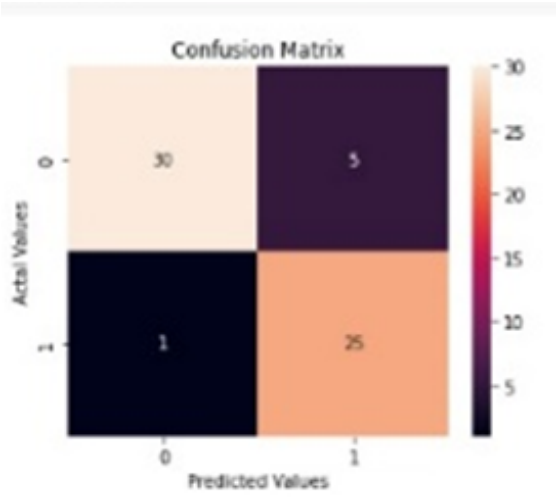
2. Root Mean Square Error (RMSE): RMSE is the quadratic scoring rule that also measures the average magnitude of the error. It is the square root of average squared difference between prediction and actual observation.

3. Mean Absolute Error (MAE): This measure represents the average absolute difference between the actual and predicted values in the dataset. It represents the average residual from the dataset.

```
RMSE:
0.31362502409359

MSE:
0.31362502409359

MAE:
0.09836065573770492

R2 SCORCE:
0.5978021978021978
```

**Classification Model:**

Confusion Matrix, Accuray Score- 0.9016 & Classification Report




CLASSIFICATION REPORT

```
[60] from sklearn.metrics import classification_report
     print(classification_report(original_classes, pred_classes))
```

```
              precision    recall  f1-score   support

         0.0       0.97      0.86      0.91        35
         1.0       0.83      0.96      0.89        26

    accuracy                           0.90        61
   macro avg       0.90      0.91      0.90        61
weighted avg       0.91      0.90      0.90        61
```

## Model Summary

```
import statsmodels.api as sm
model = sm.OLS(y_test, X_test).fit()
#view model summary
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Selling_Price   R-squared:                       0.906
Model:                            OLS   Adj. R-squared:                  0.895
Method:                 Least Squares   F-statistic:                     72.20
Date:                Mon, 14 Nov 2022   Prob (F-statistic):           7.70e-25
Time:                        12:29:52   Log-Likelihood:                -106.78
No. Observations:                  61   AIC:                             229.6
Df Residuals:                      53   BIC:                             246.4
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Present_Price     0.4370      0.044      9.868      0.000       0.348       0.526
Kms_Driven     -2.72e-05   1.24e-05     -2.200      0.032   -5.01e-05   -1.43e-05
Owner             -0.8886      1.548     -0.058      0.954      -3.178       1.400
No._of_Years      -0.2356      0.081     -2.905      0.005      -0.398      -0.073
Fuel_Type_Diesel   8.0741      1.883      8.000      0.000       6.582      10.846
Fuel_Type_Petrol   6.1682      0.909      6.787      0.000       4.345       7.991
Seller_Type_Individual   -1.2178      0.547     -2.225      0.030      -2.315      -0.120
Transmission_Manual   -1.7805      0.658     -2.702      0.007      -3.064      -0.497
==============================================================================
Omnibus:                        3.474   Durbin-Watson:                   2.357
Prob(Omnibus):                  0.209   Jarque-Bera (JB):                1.699
Skew:                           0.372   Prob(JB):                        0.428
Kurtosis:                       3.335   Cond. No.                     3.73e+05
==============================================================================
```

## 10. ADVANTAGES AND DISADVANTAGES

**Advantages:**

- Accuracy of our model is 90%.
- Prediction runs for different types of cars

**Disadvantages:**

- Accuracy can be improved.
- Prediction is done using only a few criteria.

## 11. CONCLUSION

We started with understanding the use case of machine learning in the Automotive industry and how machine learning has transformed the driving experience. We build a Random Forest Regression model to predict the resale value of a used car. Finally, we evaluated the performance of the model using the R squared score and Residual Plot.
We could have also used simpler regression algorithms like Linear Regression and Lasso Regression. Still, we need to make sure there are no outliers in the dataset before implementing them. Pair plots and scatter plots help visualize the outliers.
Then we have used a Flask application to display the predicted value to the users based on their corresponding input. This car resale value prediction can be used by the public to estimate the resale value of the car.

## 12. FUTURE SCOPE

Currently, only few features are used to predict resale value of the car. This can be extended to more features. One can also implement CNN to determine physical condition of the car from images like identifying dents, scratches etc. and thus predicting more relevant resale value of a car.
For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset. To correct for overfitting inRandom Forest, different selections of features and number of trees will be tested to check for change in performance.

## 13. APPENDIX

**Source code:**

➢ **app.py**

```
from flask import Flask, render_template, request
import jsonify
import requests
import pickle
import numpy as np
import sklearn
import datetime
from sklearn.preprocessing import StandardScaler
app = Flask(__name__)
model = pickle.load(open('car_price_prediction_model.pkl', 'rb'))


x =datetime.datetime.now()
@app.route('/',methods=['GET'])
def Home():
        return render_template('Cars.htm')


standard_to = StandardScaler()
@app.route("/predict", methods=['POST'])
def predict():
   Fuel_Type_Diesel=0
        if request.method == 'POST':
        Year = int(request.form['Year'])
      Present_Price=float(request.form['Present_Price'])
      Kms_Driven=int(request.form['Kms_Driven'])
      Kms_Driven2=np.log(Kms_Driven)
      Owner=int(request.form['Owner'])
      Fuel_Type_Petrol=request.form['Fuel_Type_Petrol']
      if(Fuel_Type_Petrol=='Petrol'):
            Fuel_Type_Petrol=1
            Fuel_Type_Diesel=0
      elif(Fuel_Type_Petrol=='Diesel'):
            Fuel_Type_Petrol=0
```

```
            Fuel_Type_Diesel=1
        else:
        Fuel_Type_Petrol=0
        Fuel_Type_Diesel=0
    Year=x.year - Year
    Seller_Type_Individual=request.form['Seller_Type_Individual']
    if(Seller_Type_Individual=='Individual'):
        Seller_Type_Individual=1
        else:
        Seller_Type_Individual=0
    Transmission_Mannual=request.form['Transmission_Mannual']
    if(Transmission_Mannual=='Mannual'):
        Transmission_Mannual=1
        else:
        Transmission_Mannual=0

prediction=model.predict([[Present_Price,Kms_Driven2,Owner,Year,Fuel_Type_Diesel,Fuel_Ty
pe_Petrol,Seller_Type_Individual,Transmission_Mannual]])
    output=round(prediction[0],2)
        if output<0:
        return render_template('Cars.htm',prediction_texts="Sorry you cannot sell this car")
        else:
        return render_template('Cars.htm',prediction_text="You Can Sell The Car for {}
lakhs.".format(output))
        else:
        return render_template('Cars.htm')
if __name__=="__main__":
    app.run(debug=True)
```

➢ **Cars.htm**

```
<html>

<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Car Resale Prediction</title>
    <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Audiowide">
    <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Sofia">
```

```
    <link rel="stylesheet"
href="https://fonts.googleapis.com/css?family=Sofia&effect=shadow-multiple">
    <style>
        body {
            background-color: khaki;
        }

        @import
url('https://fonts.googleapis.com/css2?family=Raleway:wght@900&display=swap');
        h1 {
            background: rgb(163, 57, 57);
            text-align: center;
            font-size: 70px;
            letter-spacing: 0.1em;
            -webkit-text-fill-color: transparent;
            -webkit-text-stroke-width: 3px;
            -webkit-text-stroke-color: white;
            text-shadow: 3px 3px #295bff, 6px 6px #edf033;
            font-family: "Audiowide", sans-serif;
        }

        fieldset {
            float: left;
            background-color: aliceblue;
        }

        td {
            padding-top: 3px;
            padding-bottom: 3px;
            padding-left: 8px;
            padding-right: 8px;
            font-size: 25px;
            font-family: 'Comic Sans MS', Comic Sans;
        }
```

```
h3 {
    padding-top: 20px;
}

input,
select {
    height: 25px;
    width: 200px;
    font-size: 15px;
    padding: 2px;
}

form {
    padding: 15px;
}

div {
    float: right;
    top: 100px;
}

#op {
    color: rgb(17, 0, 255);
    font-family: Copperplate, Papyrus, fantasy;
    font-size: 25px;
    text-align: center;
    font-weight: bolder;
}

.button {
    color: #fff;
    padding-top: 10px;
    padding-bottom: 10px;
```

```
        padding-left: 20px;
        padding-right: 20px;
        background-color: #38D2D2;
        background-image: radial-gradient(93% 87% at 87% 89%, rgba(0, 0, 0, 0.23) 0%, transparent 86.18%), radial-gradient(66% 66% at 26% 20%, rgba(255, 255, 255, 0.55) 0%, rgba(255, 255, 255, 0) 69.79%, rgba(255, 255, 255, 0) 100%);
        box-shadow: inset -3px -3px 9px rgba(255, 255, 255, 0.25), inset 0px 3px 9px rgba(255, 255, 255, 0.3), inset 0px 1px 1px rgba(255, 255, 255, 0.6), inset 0px -8px 36px rgba(0, 0, 0, 0.3), inset 0px 1px 5px rgba(255, 255, 255, 0.6), 2px 19px 31px rgba(0, 0, 0, 0.2);
        border-radius: 14px;
        font-weight: bold;
        font-size: 25px;
        border: 0;
        user-select: none;
        -webkit-user-select: none;
        touch-action: manipulation;
        cursor: pointer;
        float: right;
        font-family: 'Comic Sans MS', Comic Sans;
    }
  </style>

  <body>
    <h1>CAR RESALE PREDICTION</h1>
    <form action="{{ url_for('predict')}}" method="post">
      <fieldset style="width:54%;">
        <table>
          <tr>
            <td>
              <h3>YEAR</h3>
            </td>
            <td><input id="first" name="Year" type="number " placeholder="Eg. 2014" required="required"></td>
          </tr>
```

```
<tr>
  <td>
    <h3>SHOWROOM PRICE(In lakhs)</h3>
  </td>
  <td><input id="second" name="Present_Price" required="required"
placeholder="Eg. 14 (its mean 14 lakh)"></td>
</tr>
<tr>
  <td>
    <h3>KILOMETERS DRIVED</h3>
  </td>
  <td><input id="third" name="Kms_Driven" required="required"
placeholder="Eg. 24000"></td>
</tr>
<tr>
  <td>
    <h3>NO OF OWNERS PREVIOUSLY HAD THE CAR</h3>
  </td>
  <td><input id="fourth" name="Owner" required="required" placeholder="Eg.
1"></td>
</tr>
<tr>
  <td>
    <h3>FUEL TYPE</h3>
  </td>
  <td>
    <select name="Fuel_Type_Petrol" id="fuel" required="required">
    <option value="Petrol">Petrol</option>
    <option value="Diesel">Diesel</option>
    <option value="Diesel">CNG</option>
  </select>
  </td>
</tr>
<tr>
```

```
                <td>
                    <h3>DEALER OR INDIVIDUAL</h3>
                </td>
                <td>
                    <select name="Seller_Type_Individual" id="resea" required="required">
                    <option value="Dealer">Dealer</option>
                    <option value="Individual">Individual</option>
                </select>
                </td>
            </tr>
            <tr>
                <td>
                    <h3>TRANSMISSION TYPE</h3>
                </td>
                <td>
                    <select name="Transmission_Mannual" id="research" required="required">
                    <option value="Mannual">Manual Car</option>
                    <option value="Automatic">Automatic Car</option>
                </select>
                </td>
            </tr>
        </table><br>
        <button class="button" role="button" id="sub" type="submit" value="SUBMIT"
style="color:white;background-color:rgb(2, 125, 21);">SUBMIT</button>
    </fieldset>
    <div>
        <img
src="https://cdn.dribbble.com/users/515394/screenshots/5084521/media/8ff9fcc6ad5c17b43010f
6262e822ebb.gif" width="100%" height="50%">
        <br><br><br>
        <span id="op">
            <h3>{{ prediction_text }}</h3>
        </span>
    </div>
```

```
        </form>
    </body>
</head>

</html>
```

**Github source link:**
https://github.com/IBM-EPBL/IBM-Project-16120-1659607935

**Video link:**
https://vimeo.com/770488913