

**Tribhuvan University**

**Faculty of Humanities and Social Sciences**

A PROJECT REPORT ON

**Used Car Price Prediction**

**Submitted to**

**Department of Computer Application**

**Patan Multiple Campus**

**Patan Dhoka, Lalitpur**

***In partial fulfillment of the requirements for the Bachelors in Computer Application***

**Submitted by**

Himal Neupane

BCA 6th Semester

Registration No.: 6-2-22-887-2019

Under the Supervision of

**Roshan Tandukar**



**Tribhuvan University**

**Faculty of Humanities and Social Sciences**

**Patan Multiple Campus**

Patan Dhoka, Lalitpur

Bachelor in Computer Applications (BCA)

# SUPERVISOR’S RECOMMENDATION

I hereby recommend that this project prepared under my supervision by **Himal Neupane** entitled “**Used Car Price Prediction”** in the Partial Fulfillment of requirement for the degree of Bachelor in Computer Application is recommended for that final evaluation.

Roshan Tandukar

Project Supervisor

BCA Department

Patan Multiple Campus

# ABSTRACT

With the rise in social standards, cars are becoming an everyday necessity for many people, driving up the demand for used vehicles. This program uses the Random Forest Algorithm to predict used car prices by analyzing various factors such as the car’s brand, model, age, mileage, and engine displacement, max power, kilometers driven. It also considers additional features like the type of seller, fuel type, transmission, and seating capacity to estimate a fair market value for each car. The goal of this system is to provide a reliable price estimate that benefits both buyers and sellers. By using the Random Forest Algorithm, which is excellent at handling complex data and capturing intricate relationships between different factors and car prices, the program aims to make the used car market more transparent and efficient. This helps everyone involved make more informed and fair decisions when buying or selling a used car.

**Keywords**: Random Forest Algorithm, Used Car Price, Vehicle Age, Mileage, Engine ,



**Tribhuvan University**

**Faculty of Humanities and Social Sciences**

**Patan Multiple Campus**

Patan Dhoka, Lalitpur

Bachelor in Computer Applications (BCA)

# LETTER OF APPROVAL

This is certify that this project prepared by **Himal Neupane** entitled “**Used Car Price Prediction”** in the Partial Fulfillment of requirement for the degree of Bachelor in Computer Application has been evaluated. In our opinion it is satisfactory in the scope and quality as a project for the required degree.

Roshan Tandukar

Supervisor

BCA Department

Patan Multiple Campus

Roshan Tandukar

Program Coordinator

Patan Multiple Campus

Patan Dhoka, Lalitpur

# ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to our supervisor **Mr. Roshan Tandukar** who gave me the golden opportunity to do this wonderful project on the topic of **Used Car Price Prediction**, which also helped us in doing a lot of research.

I would like to express my special thanks of gratitude to our Campus Chief Dr**. Laxman Singh Kunwar** who gave me permission for doing this Project.

I would like to express my special gratitude and thanks to our BCA Program Coordinator **Mr. Roshan Tandukar** for his support and help for our personnel development and mainly for the completion of this Project.

I thank Patan Multiple Campus a lot for their direction and continuous supervision, as well as for providing the knowledge that I required to understand the project and for assisting with seeing it through to its completion..

I would also like to thank my friends who helped me a lot in finalizing this project within the limited time frame.

Yours sincerely,

Himal Neupane

**TABLE OF CONTENTS**

[SUPERVISOR’S RECOMMENDATION ii](#_Toc175251926)

[ABSTRACT iii](#_Toc175251927)

[LETTER OF APPROVAL iv](#_Toc175251928)

[ACKNOWLEDGEMENT v](#_Toc175251929)

[CHAPTER 1: INTRODUCTION 1](#_Toc175251930)

[1.1 Introduction 1](#_Toc175251931)

[1.2.Problem Statement 1](#_Toc175251932)

[1.3.Objectives 2](#_Toc175251933)

[1.4 Scope and Limitation 2](#_Toc175251934)

[1.5 Development Methodology 2](#_Toc175251935)

[1.6 Report Organization 3](#_Toc175251936)

[CHAPTER 2: BACKGROUND STUDY AND LITERATURE REVIEW 4](#_Toc175251937)

[2.1 Background Study 4](#_Toc175251938)

[2.1.1 Fundamental Theories and General Concepts 4](#_Toc175251939)

[2.2 Literature Review 5](#_Toc175251940)

[CHAPTER 3: SYSTEM ANALYSIS AND DESIGN 6](#_Toc175251941)

[3.1 System Analysis 6](#_Toc175251942)

[3.1.1 Requirement Analysis 6](#_Toc175251943)

[3.1.2 Feasibility Analysis 8](#_Toc175251944)

[3.1.3 Data modelling: ER Diagram 9](#_Toc175251945)

[3.1.4 Process Modelling: DFD 9](#_Toc175251946)

[3.2 System Design 12](#_Toc175251947)

[3.2.1 Architectural design 12](#_Toc175251948)

[3.2.2 Database Schema design 13](#_Toc175251949)

[3.2.3 Interface design (UI/UX) 13](#_Toc175251950)

[3.2.4 Physical DFD 16](#_Toc175251951)

[3.3 Algorithm details 16](#_Toc175251952)

[CHAPTER 4: IMPLEMENTATION AND TESTING 19](#_Toc175251953)

[4.1 Implementation 19](#_Toc175251954)

[4.1.1Tools Used: 19](#_Toc175251955)

[4.1.2 Implementation details of modules 19](#_Toc175251956)

[5. APPENDICES 20](#_Toc175251957)

[6. REFRENCES 20](#_Toc175251958)

**LIST OF ABBREVIATIONS**

CSS Cascading Style sheets

DFD Data Flow Diagram

ERD Entity Relationship diagram

HTML Hypertext Markup Language

JS JavaScript

UI User Interface

UX User Experience

**LIST OF FIGURES**

Figure 1.1 Waterfall Model 3

Figure 3.1 Use case diagram 7

Figure 3.2 ER-Diagram 9

Figure 3.3 Level 0 DFD 10

Figure 3.4. Level 1 DFD 11

Figure 3.5 Architectural design 12

Figure 3.6 Database Schema design 13

Figure 3.8 Interface design (UI/UX) 13

Figure 3.9 Physical DFD 16

Figure 5.1 Appendices 20

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction

The increasing demand for used cars makes used car price prediction an important factor in the car industry. Determining an affordable cost for a used car becomes more challenging as more people search for affordable options. A wide range of components, like make, model, year, mileage, condition, and market trends, must be examined. Because of this complexity, accurate techniques needs to be created to help sellers and buyers make smart decisions.

The system lets users enter information about the car they want to buy or sell, such as the brand, model, mileage, transmission type, fuel type, kilometers driven . seats, max power and vehicle age. Using Random Forest algorithm we can analyzes those data and compares it to a vast database of past history of cars sold and bought and market trends. Theoutcome of this algorithm is reliable price prediction that gives the current market value of the car.

In essence, the Used car Price Prediction System is a web based system that provides users with the information they need to make safe and intelligent decisions in the used cars market. The program simplify the process of estimating the true worth of a previously owned car using new technology and information data analysis, enabling both buyers and sellers to perform transactions with greater confidence .

## 1.2.Problem Statement

In today's quickly changing car sector, both buyers and sellers suffer from difficulties checking the true market value of used cars. The lack of information and inconsistent pricing can result in great price variations across identical cars. Buyers may overpay for vehicles, while sellers may undersell their cars, causing unsatisfactory selling or buying..

To address these problems, the program aims to develop a trustworthy and precise used cars price predict model. The program will give users pricing estimates using data such as make, model, year of manufacturing, mileage, condition, and market trends. This would allow users to make knowledgeable decisions, insure fair purchases, and optimize marketplace accuracy**.**

## 1.3.Objectives

The main objectives of Used Car Price Prediction System are as below:

* To predict the price of used cars using Random Forest Algorithm
* To have Exportable History for logged-in users.

## 1.4 Scope and Limitation

**i. Scope:**

Factors like brand, model , Age of the vehicle is, owned by , kilometers driven ,transmission type ,mileage, fuel type, and its engine size plays a role in determining car prices.

The system assist buyers and sellers find a good price for used cars which in turn lead them to make proper decision.

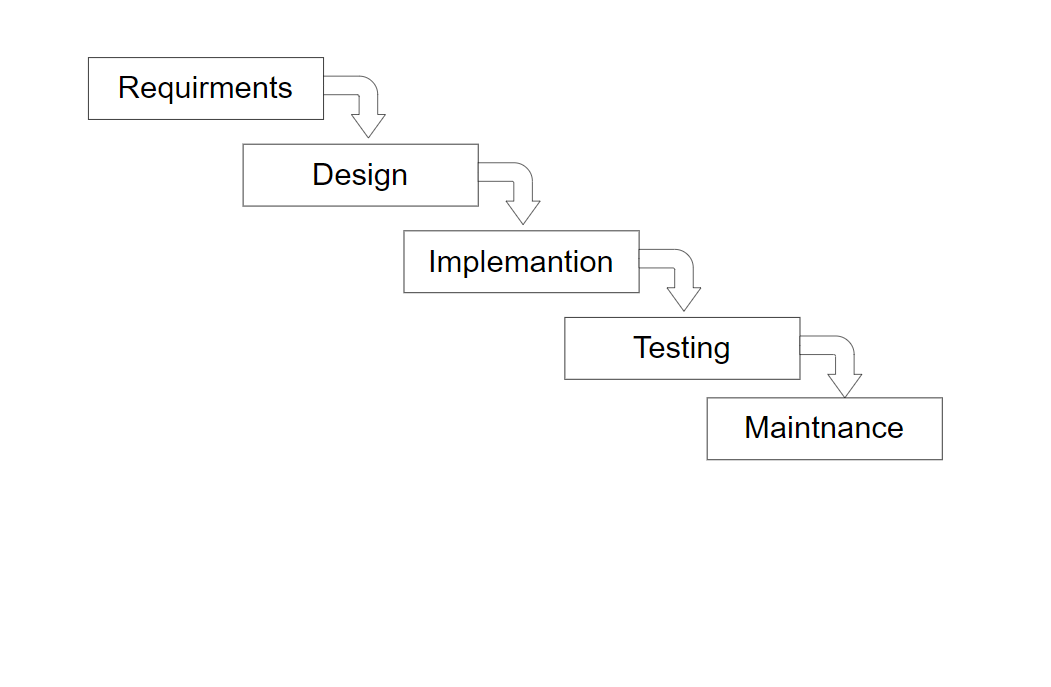
**ii. Limitations:**

Since we do not have the features available for each of these cars, our predictions are only as good as the data and may be even very far from actual prices.

These prices do not apply to the specific vehicle of an owner. External factors such as fuel-prices, market-trends or economic-factors are also excluded that affect carprices.

## 1.5 Development Methodology

The process flow for Used Car Price Prediction includes analysis of the requirements, design, implementation, testing, and maintenance. During the requirement analysis process, every functional and non-functional requirement is examined and the system is then developed to meet the requirements. The system is integrated and tested after the design phase is followed by the coding and development phase. The system is installed if the testing is successful; if not, some maintenance is carried out prior to the system being used.



**Figure 1.1 Waterfall Model**

## 1.6 Report Organization

The purpose of this report is to set forth the project as a whole. It starts with Chapter 1: Introduction, which gives the background of the project, problems it is intended to solve, its objectives, its scope and limitations, method used for development and how the report is arranged. Chapter 2: Background Study and Literature Review defines basic theories, concepts and terms relating to the subject matter while exploring similar projects and previous researches done by others. Chapter 3: System Analysis and Design highlights how the system was analyzed and designed. This encompasses what should be done, how it can be done, structure of data and process flowcharting. The chapter also deals with system’s architecture, database design and user interface design. Finally in chapter 4; Implementation and Testing tells us how the system was developed as well as tested including tools/platforms used and deployment setup

# CHAPTER 2: BACKGROUND STUDY AND LITERATURE REVIEW

## 2.1 Background Study

Buying a used car is an big financial decision for individuals and families in the present automotive industry. But when we consider all the things that influence price. Traditional prediction methods like inspecting up vehicle prices or negotiating with dealerships can be time-consuming.

Traditionally, calculating the value of a used car has instead seen users copmare through multiple sources to compare different models and take into account factors such as vehicle condition and trends in the current market. It can be a labor-intensive procedure that lacks the accuracy required to make well-informed decisions, which translates into limited data on whether or not user is paying market rate.

### 2.1.1 Fundamental Theories and General Concepts

Predictive Modeling: Predictive modeling is the use of statistical techniques and algorithms to make predictions or forecasts based on historical data. Example if predictive modeling predicts car price, the aim is to appraise how much an automobile would sell for using its brand, model, age, mileage among others. An instance of this is regression analysis that models the relationship between selling price and other variables

Decision Tree: A decision tree is a tool used to make decisions by breaking down a complex problem into a series of simple yes-or-no questions. Decision trees are easy to understand and visualize, making them helpful for tasks like classifying items or predicting outcomes. However, if a decision tree becomes too detailed and complicated, it might not perform as well, which is why it’s often combined with other methods, like Random Forests, to improve accuracy and handle more complex situations.

Random Forest Algorithm: Random forest is a strategy where several decision trees are created to make forecasts. It doesn’t rely on one single tree that can be mistaken but rather combines the results of many trees to give a more precise approximation. It’s best suited for handling voluminous data with numerous features that bear complex patterns among them. Therefore, it is suitable for estimation of used car price or any other case where relationship between variables is not linear

## 2.2 Literature Review

There are many similar applications that has been developed or in developing process available in web. Nowadays as the net income of people increase they tend to go for a used car for first hand experience of car for themselves and decide whether it is useful for them or not.

Cash On Wheel is an online platform dedicated to buying and selling cars at competitive prices. It provides detailed information on each vehicle, including specifications, pricing, and seller contact details. In this website the user requires to enter their phone number to see the predicted price so if any user want to check price they have to wait for the contact by the company using the website. Furthermore, they have list of few cars as well as dummy cars which ,ay have been used in testing phase. [1]

CarWale's used car price prediction tool provides a comprehensive platform for users to estimate the market value of their used cars. Users provide details like make, model, year, city they are now and Kilometers driven. It also has feature to select if the car is first hand second hand or more. This allows user to know more but its only available in major cities of India. [2]

Car Hamro is also a website focused on the sale and purchase of used cars. The website provides users a thorough listing of different car brands and models, along with complete details about each one, such as price, specs, and seller contact details. Likewise it is a marketplace so it doesn’t provide any assistance with price prediction so first drivers for cars can get hesitant about it as it may not be fair price according to market also the sellers also may get confused about fair price of their car as it doesn’t have any prediction of car value..[3]

Ramro Motors provides full details on various kinds of used cars, like model year, price, fuel and type. It provides a wide choice of alternatives for potential buyers by providing to various car manufacturers and types. It is also a marketplace so it doesn’t provide any estimates for used car prediction.[4]

# CHAPTER 3: SYSTEM ANALYSIS AND DESIGN

## 3.1 System Analysis

3.1.1 Requirement AnalysisThe requirements are to be collected prior to beginning projects’ development life cycle. Both functional and non-functional requirements of the system have been researched in order to build and create it.

#### i. Functional Requirements

**For Admin**

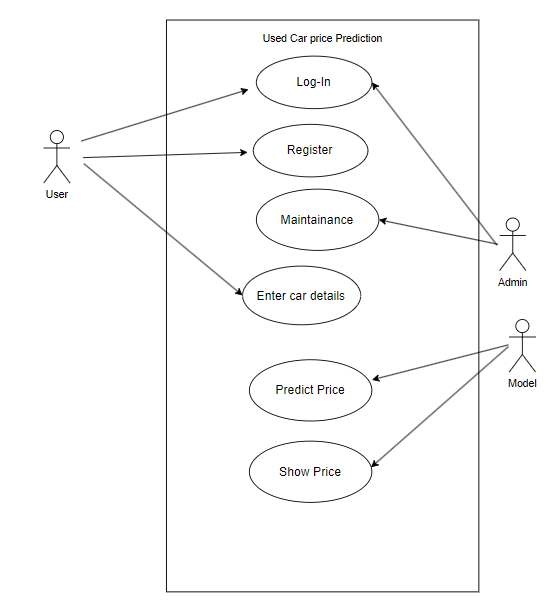
* System should allow admin to login and logout from the system
* System should allow admin to add or delete users.

**For User**

* System should allow user to register, login and logout
* System should allow user to predict the price

**Usecase Diagram**

In Used car price prediction, the use case diagram consists of user where user can register and login using own credentials. Likewise admin is allowed to update users, view the history and maintain the system.



**Figure 3. 1 Use case diagram**

#### ii. Non Functional Requirement

Different non-functional requirement have been studied and identified and are listed as below:

**Security:** Admin representative will be able to log into the system and have access to the system but access to have various subsystems will be protected by the user login screen that requires a username and password.

**Availability**: The system will be available for 24 hours service as Used car price prediction

**Reliability**: The system will be reliable as it will perform function and run without a failure, and it has to be reliable due to importance of data and damages that can be caused by incorrect prediction.

### 3.1.2 Feasibility Analysis

#### i. Technical Feasibility Analysis

The system will be technically feasible as the requirement for the development of the system is easily accessible. The basic programming language which is appropriate for project is accessible and the set of libraries required for project is capable of achieving the result that I am aiming for. For system creation and maintenance, every resource now in existence may be employed.

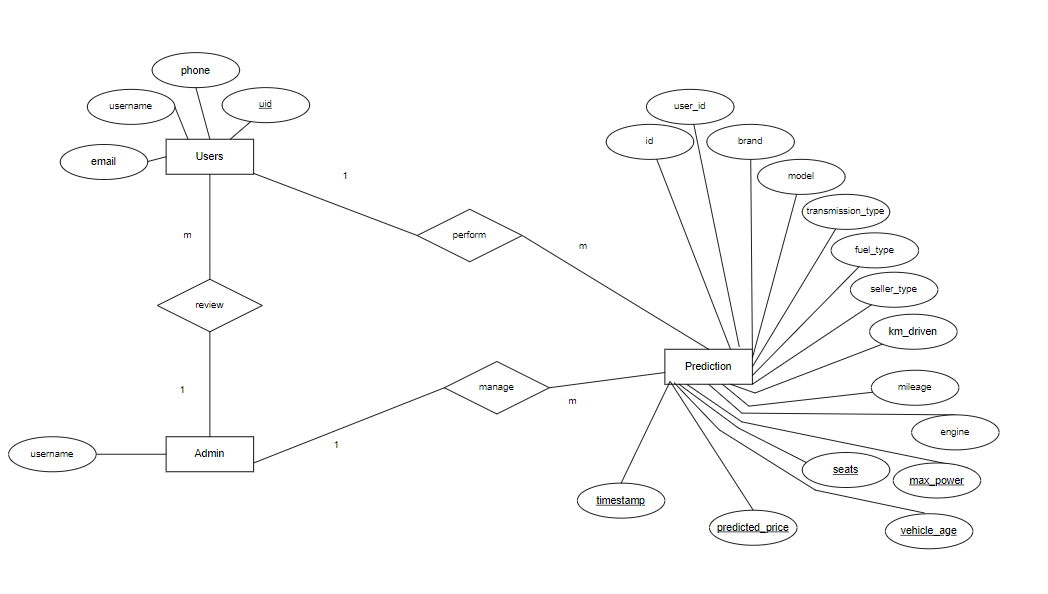
#### ii.Operational Feasibility Analysis

The system will be easy to operate with the basic knowledge of using the internet and computing devices and well trained manpower is not necessary. User can also easily access the system as it is user friendly in many aspects with good User Interface. This system include all the requirements used for Used car price prediction.

#### iii) Economic Feasibility Analysis

The proposed system is economically viable and cost-effective, utilizing open-source tools and resources. There will be no need for additional hardware or software deployment post-completion, leveraging existing infrastructure effectively.

### 3.1.3 Data modelling: ER Diagram



**Figure 3. 2 ER-Diagram**

### 3.1.4 Process Modelling: DFD

Data Flow Diagram consists of two levels of DFD which are context diagram and level 1 DFD. This are used to make the system.

**Level 0: Context DFD**

In Level 0 DFD, login request, prediction, registration is input for LLPS and user request for login and register and admin request for login. User will perform prediction by giving in some specification.

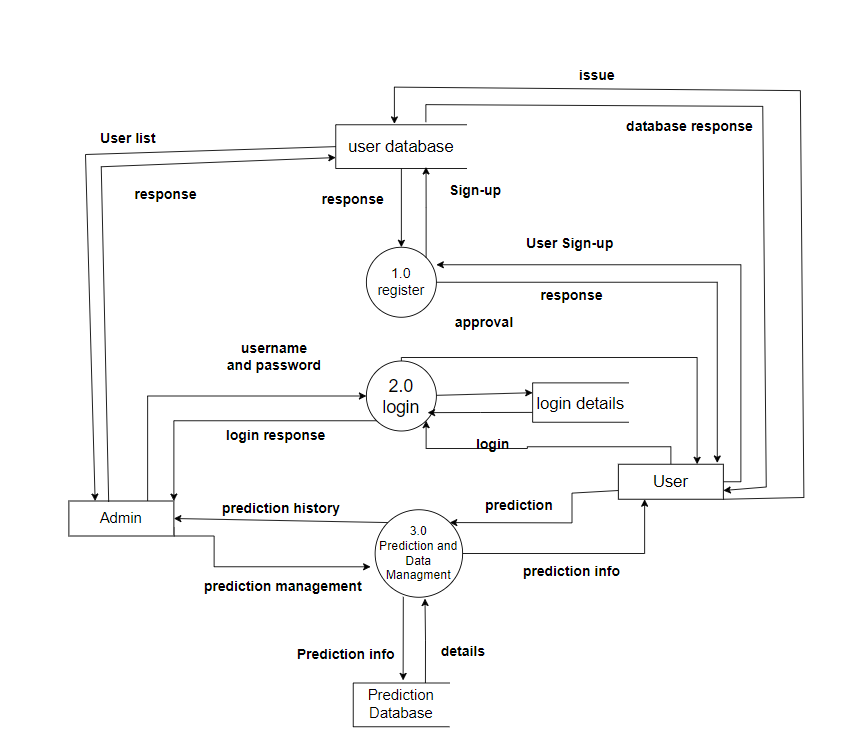
A diagram of used car price prediction

Description automatically generated

**Figure 3.3 Level 0 DFD**

**Level 1 DFD:**

In level 1 DFD there are three processes where login is responsible for login of admin and users. Prediction is responsible for user prediction. Registration is responsible for user registration. There are three entities Admin, user and system and two data stores which are users and predictions.



**Figure : 3.4 Level 1 DFD**

## 3.2 System Design

System design involves defining the architecture, components, and interfaces of a system to meet specified requirements. It includes the planning of both hardware and software elements to ensure they work together efficiently and effectively.

### 3.2.1 Architectural design

In architectural design, used car price Prediction system user interacts with User Interface which is connected to web server which is linked to machine Learning model which sends predicted data to database and vice-versa.

A diagram of a computer server and machine learning

Description automatically generated

**Figure 3. 5Architectural design**

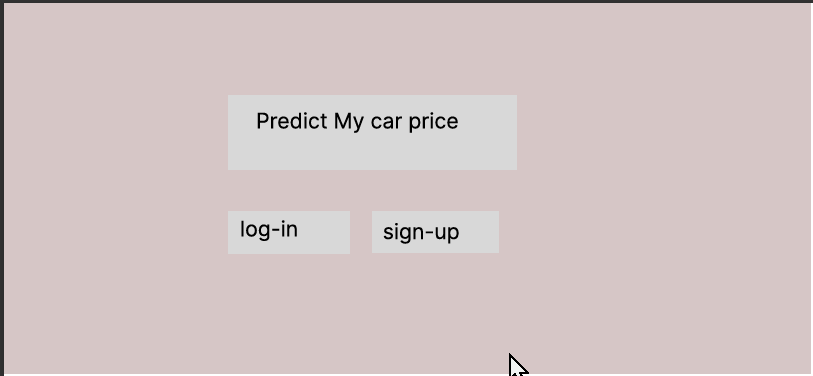
### 3.2.2 Database Schema design

A screenshot of a computer

Description automatically generated

**Figure 3. 6 Database Schema design**

### 3.2.3 Interface design (UI/UX)



A screenshot of a pink box

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A white background with black text

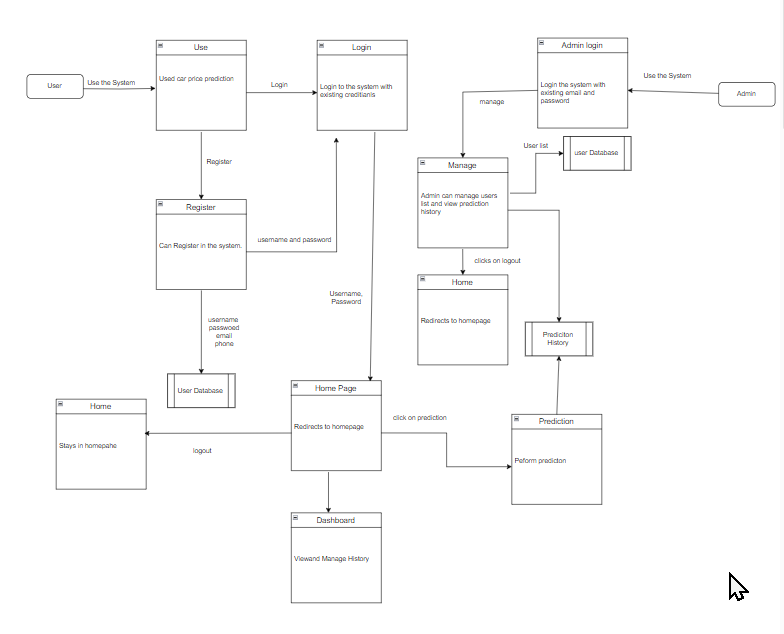
Description automatically generated

A screenshot of a computer screen

Description automatically generated

**Figure 3. 7 Interface design (UI/UX)**

### 3.2.4 Physical DFD



**Figure 3. 3 Physical DFD**

## 3.3 Algorithm details

Decision Tree and Random Forest Algorithms

Decision Tree: A decision tree splits the data into subsets based on the values of the features to minimize prediction error. Each node in the tree represents a decision based on a feature's value, leading to branches that further split the data until a stopping criterion is met. The decision tree algorithm uses the Mean Squared Error (MSE) formula to find the best split:

MSE = (Var(left\_y) \* len(left\_y) + Var(right\_y) \* len(right\_y)) / len(y)

where `Var` denotes the variance, `len` is the number of samples in each subset, and `y` represents the target values. The algorithm evaluates all possible splits by iterating over each feature and possible split values, calculating the MSE for each split, and selecting the one with the lowest MSE to ensure the most informative and optimal division of the data.

Random Forest: A random forest enhances the decision tree model by creating an ensemble of trees, each trained on a different bootstrap sample of the training data. The key advantage of the random forest is its ability to reduce overfitting and improve generalization. During training, each tree is constructed using a random subset of features and samples, which introduces diversity among the trees. For prediction, the random forest aggregates the predictions from all trees by averaging them, providing a robust and stable prediction:

Predicted Value = (1 / N) \* sum(Prediction\_i for i in range(N))

where `N` is the number of trees in the forest, and `Prediction\_i` is the prediction made by the i-th tree. This averaging process helps in reducing the variance and improving the overall performance of the model.

**Pseudo Code**

N = number\_of\_trees

M = number\_of\_features\_per\_split

for i in 1 to N:

sample = bootstrap\_sample(dataset)

tree = build\_tree(sample, M)

forest.append(tree)

function build\_tree(data, M):

if stopping\_condition\_met:

return leaf\_node

features = random\_sample(all\_features, M)

split = find\_best\_split(data, features)

left, right = split\_data(data, split)

return Node(split, build\_tree(left, M), build\_tree(right, M))

function predict(input):

votes = [tree.predict(input) for tree in forest]

return majority\_vote(votes)

# CHAPTER 4: IMPLEMENTATION AND TESTING

## 4.1 Implementation

### 4.1.1Tools Used:

* **CASE Tools:** Figma, Canva and Draw.io were used for creating and managing deployment diagrams. These tools provide robust features for designing and visualizing system architectures.
* **Programming Languages:** The application is built using Python for the backend with Flask for handling web requests and managing server-side logic, HTML and CSS for the frontend, and JavaScript for dynamic functionalities. These languages are essential for developing and integrating various system components.
* **Database Platforms:** SQLite is used as the database platform. It is a lightweight, file-based database that supports SQL queries and efficiently handles data storage needs.
* **Dataset:** The data used for the application was sourced from Kaggle, a well-known platform for datasets and machine learning competitions. This dataset includes features relevant to predicting used car prices and was used to train and evaluate the machine learning model. [5]

### 4.1.2 Implementation details of modules

I. Decision Tree Class

* \_\_init\_\_(self, max\_depth=None): Sets up the decision tree, optionally limiting its depth.
* fit(self, X, y): Teaches the tree to make predictions by repeatedly splitting the data to find the best splits.
* predict(self, X): Uses the trained tree to make predictions based on the input data.

II. Random Forest Class

* \_\_init\_\_(self, n\_estimators=100, max\_depth=10): Sets up a collection of decision trees with a specified number of trees and optional depth limit.
* fit(self, X, y): Trains each tree on different samples of the data and combines their results.
* predict(self, X): Averages the predictions from all trees to get the final result.

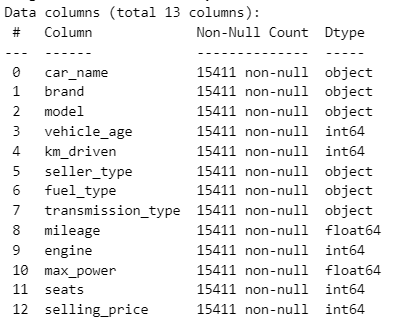
III. Model Loading and Prediction

* Model Loading: Loads pre-trained models using joblib.
* Prediction: Prepares the input data, applies the trained Random Forest model, and returns the predictions.

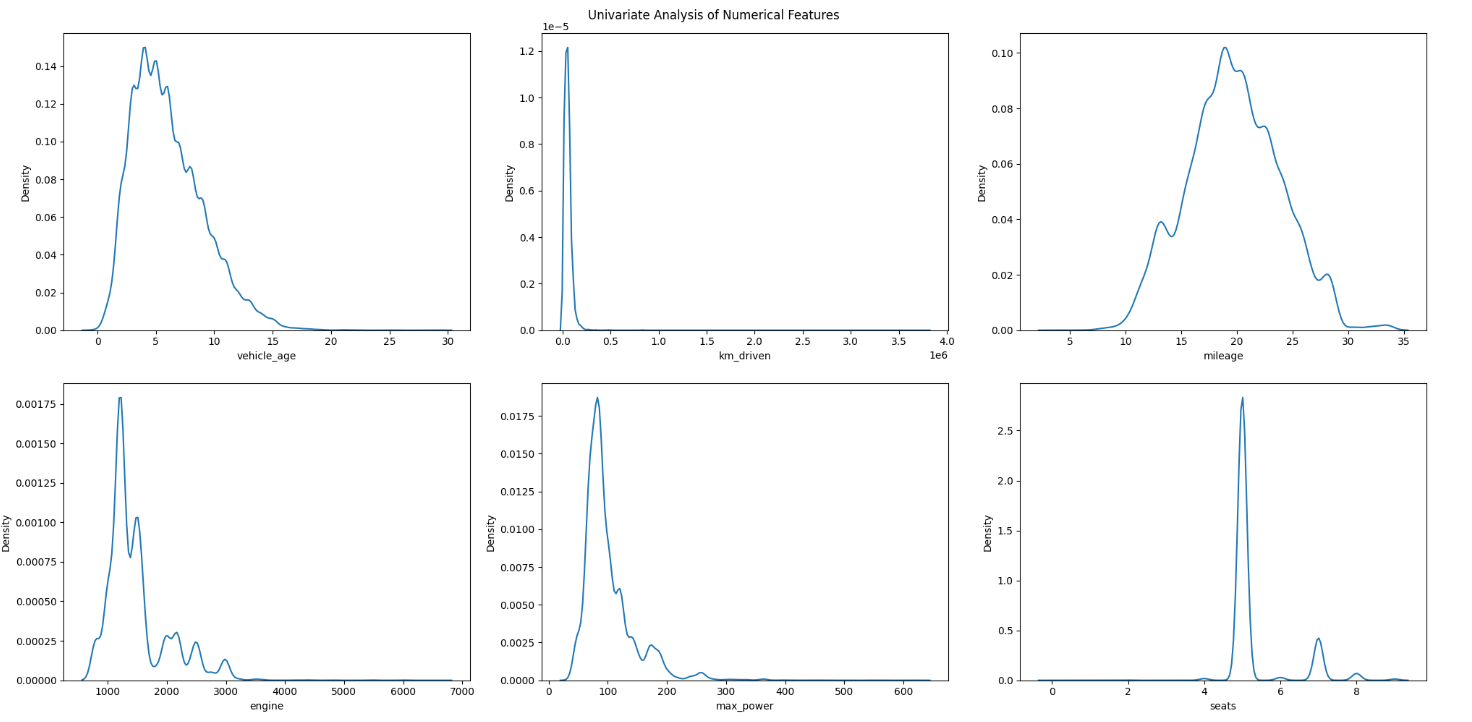
# 5. APPENDICES

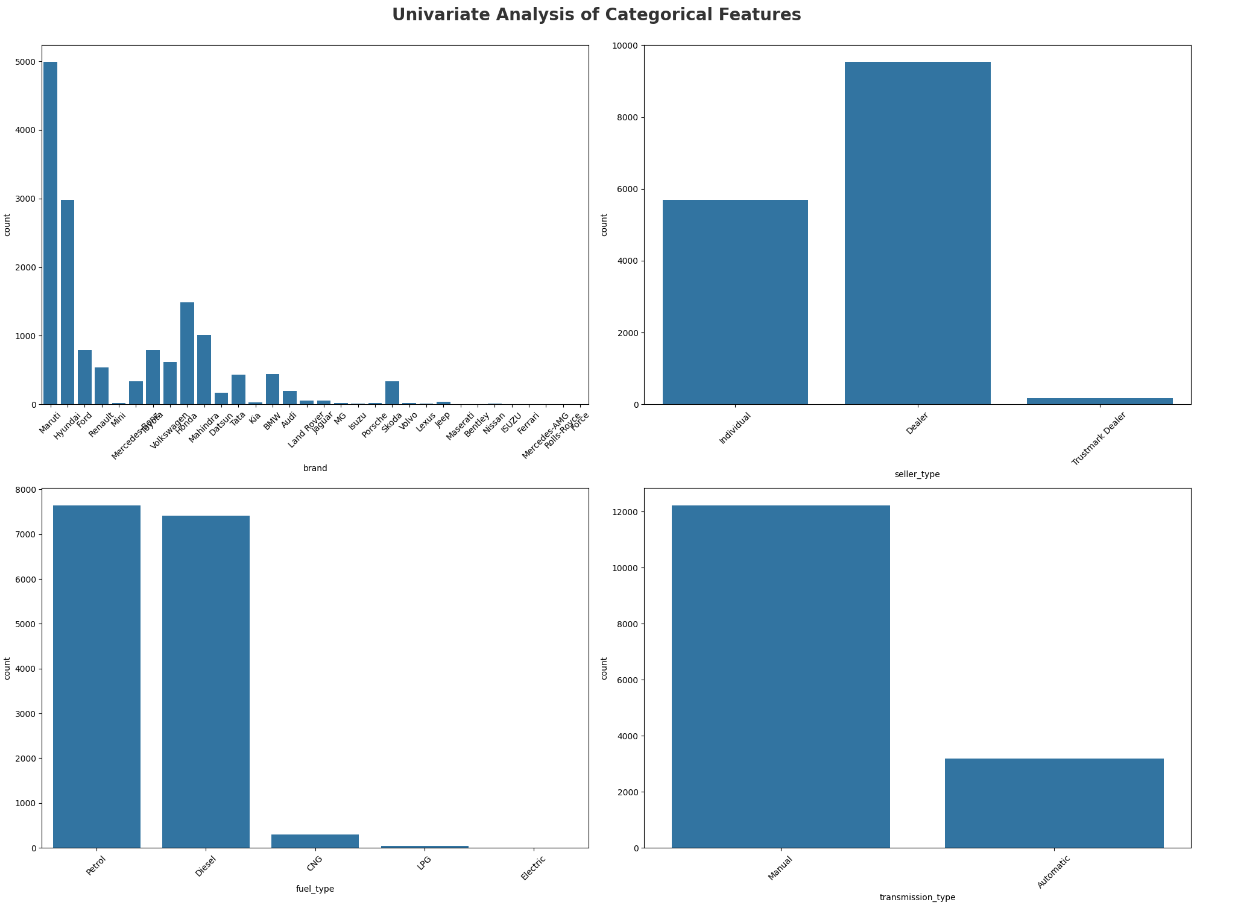
A screenshot of a computer

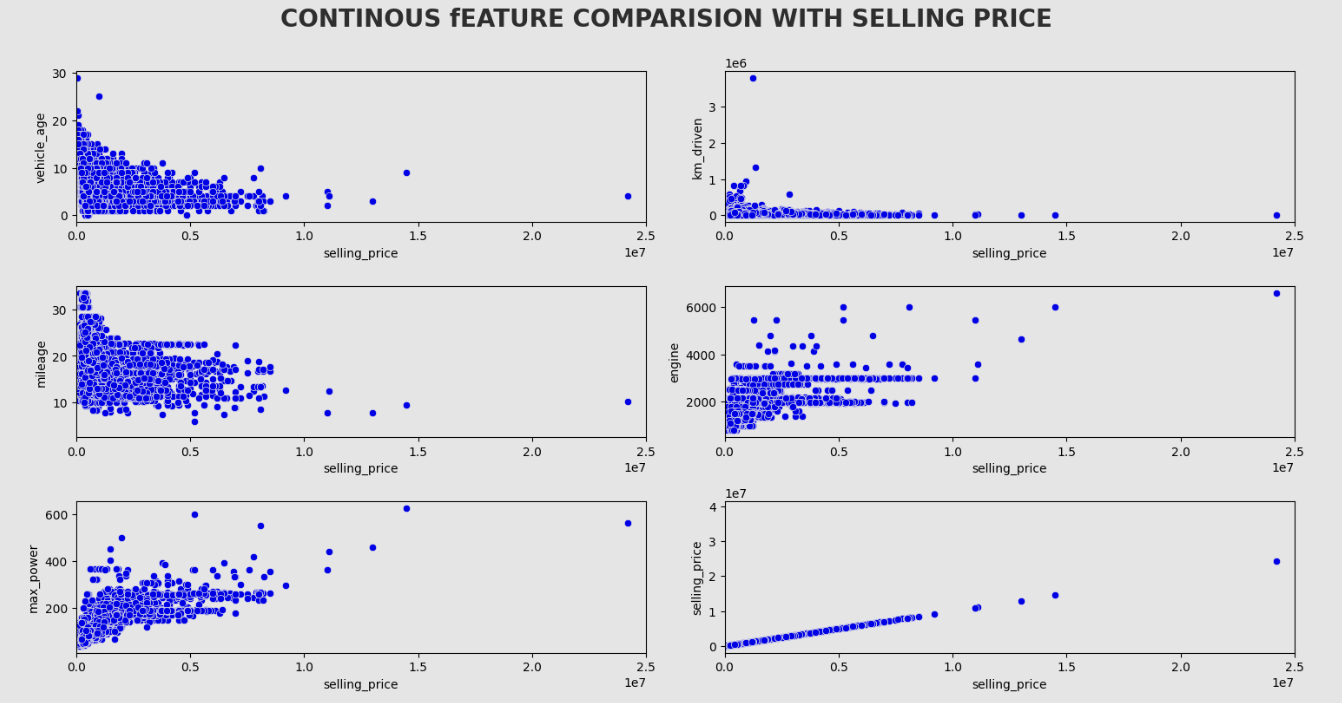
Description automatically generated



A graph of a price

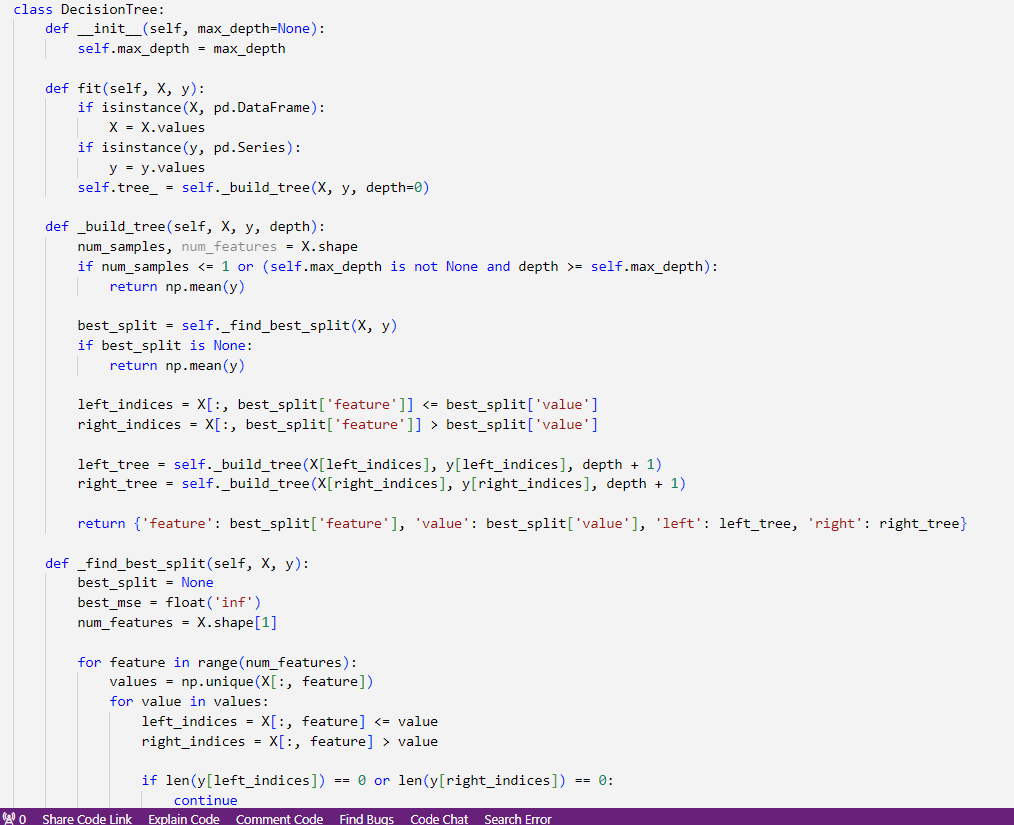
Description automatically generated with medium confidence





A computer code with black text

Description automatically generated with medium confidence

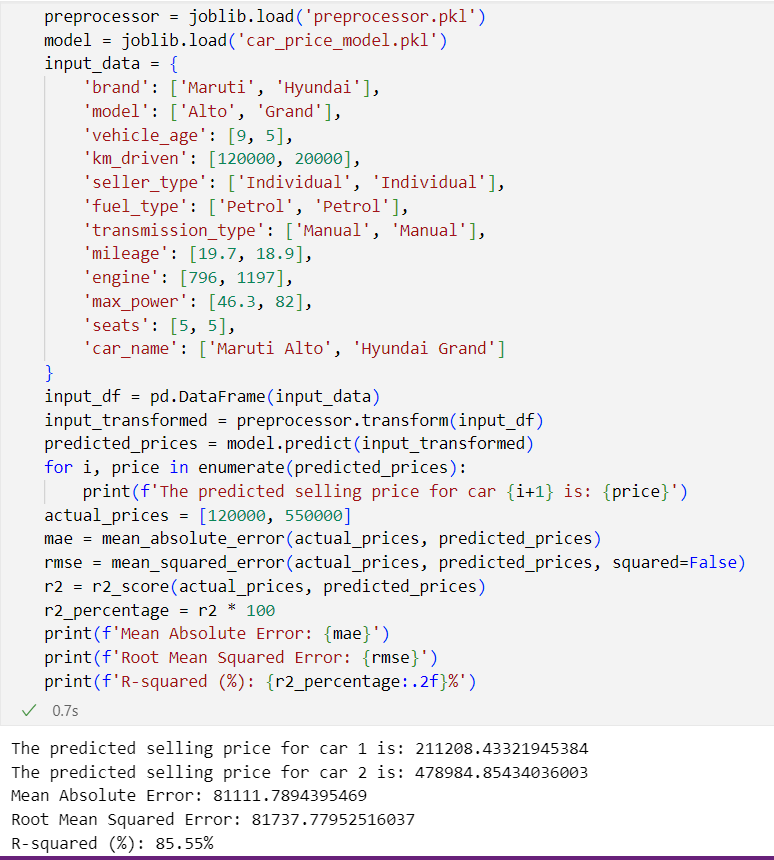


A screenshot of a computer program

Description automatically generated

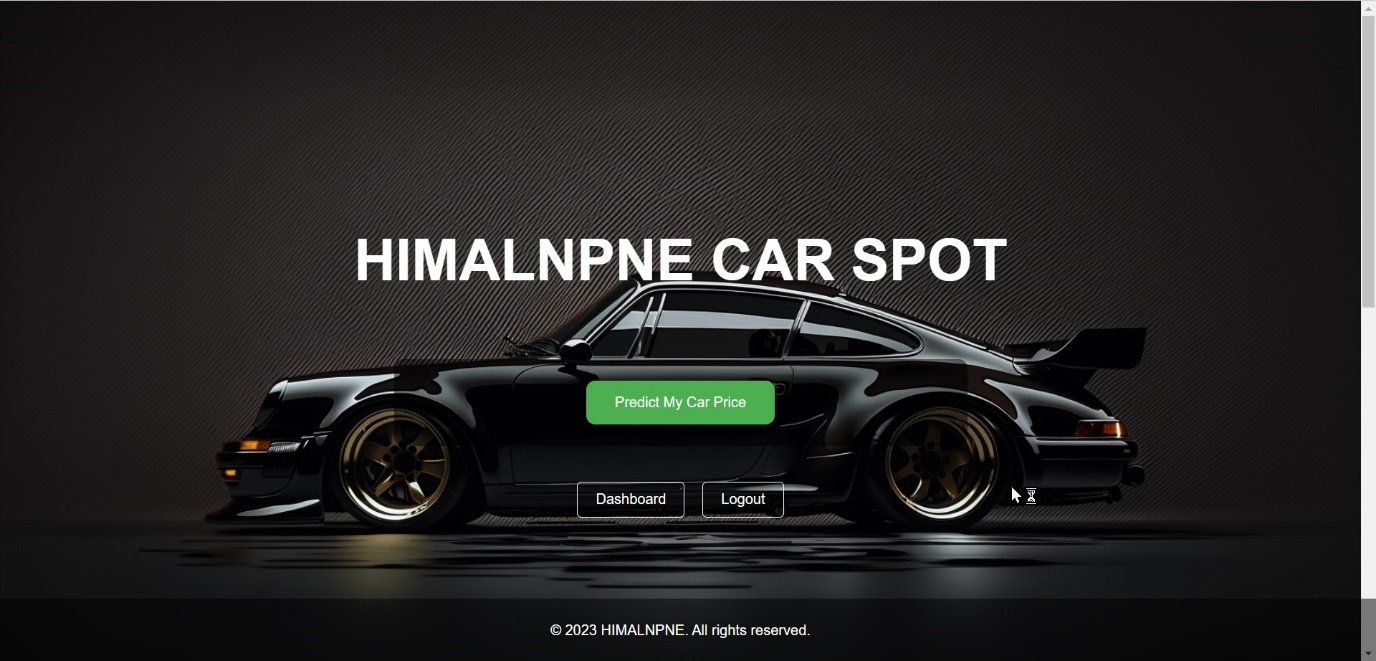
A screenshot of a computer program

Description automatically generated



A black car with a green sign

Description automatically generated



A screenshot of a car price prediction form

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a login form

Description automatically generatedA screenshot of a login screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

# 6. REFRENCES

[1] "Cash On Wheel :: Used car in Nepal price," Cash On Wheel. [Online]. Available: https://cashonwheel.com/. [Accessed: 22-Aug-2024].

[2] "Used Car Price prediction," CarWale. [Online]. Available: https://www.carwale.com/used/carvaluation/. [Accessed: 22-Aug-2024].

[3] "Used Cars in Nepal," Car Hamro. [Online]. Available: https://carhamro.com/ [Accessed: 22-Aug-2024].

[4] "Used car price in Nepal," Ramro Motors. [Online]. Available: https://ramromotors.com.np/search-for-used-vehicles/. [Accessed: 22-Aug-2024].

[5] M. Kumar, "CarDekho Used Car Data," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/manishkr1754/cardekho-used-car-data/. [Accessed: 22-Aug-2024].