A Mini Project Report on

Car Price Prediction

T.E. - I.T Engineering

Submitted By

Shivam Gupta	**20104110**
Yashab Mahimi	**20104026**
Meet Bohra	**20104057**
Kaif Khan	**20104128**

Under The Guidance Of

Prof. Geetanjali Kalme

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year : 2022-23

CERTIFICATE

This to certify that the Mini Project report on **Car Price Prediction** has been submitted by **Shivam Gupta (20104110), Yashab Mahimi (20104026), Meet Bohra (20104057), and Kaif Khan (20104128)** who are a Bonafide students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2022-2023** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Geetanjali Kalme
Guide

Dr. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)
1.
2.

Place:A.P.Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof. Geetanjali Kalme** Expressing gratitude towards our HoD, **Dr. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher Ms. **Charul Singh** who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

ABSTRACT

In recent years, while the scale of the used car market has been expanding, the price evaluation system of country's second-hand car market has exposed the problem that it does not meet the market demand. Accurate used car price prediction can help people make correct decisions and avoid the wanton price tag of used cars in the market as much as possible. This project uses the random forest algorithm to predict the price of used cars and compares and analyzes the prediction results.

In this study, we explore the use of the random forest algorithm for car price prediction. We begin by collecting a large dataset of car prices and their associated features, including the car's make, model, year, mileage, and other relevant attributes. We then preprocess and clean the data, removing any outliers or missing values. Next, we use the random forest algorithm to train a regression model on the dataset, using a subset of the features as input variables.

Our results show that the random forest algorithm is a highly effective method for car price prediction, achieving high levels of accuracy and outperforming other commonly used regression algorithms. Our analysis also highlights the importance of certain features, such as the car's age and mileage, in predicting its market value. Overall, our study provides valuable insights into the use of machine learning algorithms for car price prediction and offers practical applications for the automotive industry.

TABLE OF CONTENTS

Chapter 1

Introduction

Car price prediction is a crucial problem in the automotive industry, as it plays a significant role in determining the market value of a car. Accurately predicting the price of a car can help both buyers and sellers make informed decisions, leading to more efficient transactions and a more transparent market.

In recent years, there has been a growing interest in using machine learning algorithms to predict car prices. These algorithms can analyze a wide range of features, such as the make and model of the car, its age, mileage, and other relevant attributes, to provide a more accurate estimation of its value. Additionally, machine learning algorithms can be trained on large datasets of historical car prices, allowing them to identify patterns and trends in the market that may not be immediately apparent to humans.

One popular machine learning algorithm used for car price prediction is the random forest algorithm. This algorithm is a type of decision tree that combines multiple decision trees to reduce overfitting and improve prediction accuracy. Random forests are highly effective for regression tasks, making them well-suited for car price prediction.

In this project, we explore the use of the random forest algorithm for car price prediction. We aim to develop a model that can accurately predict the market value of a car based on its features. To achieve this, we collect a large dataset of car prices and their associated attributes and use this data to train and evaluate our model. We also perform feature importance analysis to identify which attributes are most important in predicting car prices.

1.1. **Purpose :**

The purpose of car price prediction is to provide accurate estimates of the market value of a car based on its features. This information is valuable to both buyers and sellers, as it helps them make informed decisions about buying or selling a car at a fair price. By accurately predicting the value of a car, sellers can avoid pricing their car too high or

too low, which can result in lost sales or lower profits. Similarly, buyers can avoid overpaying for a car or missing out on a good deal.

Car price prediction also has practical applications for the automotive industry. Dealerships and car manufacturers can use these predictions to better understand the market and adjust their pricing strategies accordingly. They can also use these predictions to identify which features are most valuable to customers and adjust their production and marketing strategies to meet these demands.

1.2. Problem Statement:

The problem statement for car price prediction is to accurately estimate the market value of a car based on its features.

One of the key challenges in car price prediction is identifying the most important features that influence car prices. This requires careful analysis of large datasets of historical car prices and the use of advanced machine learning algorithms to identify patterns and trends in the market. Another challenge is developing a model that is robust to outliers and can provide accurate predictions even when the data is noisy or incomplete.

Ultimately, the problem statement for car price prediction is to provide valuable insights and information to stakeholders in the automotive industry. By developing accurate prediction models and identifying the most important features that influence car prices, we can help buyers and sellers make informed decisions, improve pricing strategies, and drive innovation in the automotive industry.

1.3. Objectives :

The objectives of car price prediction are as follows:

- Develop an accurate prediction model: The primary objective of car price prediction is to develop a prediction model that can accurately estimate the market value of a car based on its features. This model should be able to analyze a wide range of attributes and provide reliable estimates of car prices.

- Identify important features: In addition to developing an accurate prediction model, car price prediction also aims to identify the most important features that influence car prices. By understanding which features have the greatest impact on car prices, we can develop better pricing strategies and production plans for the automotive industry.

- Compare different algorithms: Car price prediction involves testing and comparing different machine learning algorithms to determine which ones are most effective for this task. By comparing algorithms, we can identify which ones provide the most accurate predictions and improve our understanding of the strengths and limitations of each method.

- Optimize prediction performance: The objective of car price prediction is to optimize prediction performance by fine-tuning the prediction model and selecting the most important features. This process involves adjusting the parameters of the machine learning algorithm and identifying which features should be included in the model to improve its accuracy.

1.4. Scope:

The scope of car price prediction includes the development of machine learning models that can accurately estimate the market value of a car based on its features. This involves analyzing a wide range of attributes, such as the make and model of the car, its age, mileage, and other relevant factors, to provide reliable estimates of car prices.The scope of car price prediction also includes the identification of the most important features that influence car prices. By understanding which features have the greatest impact on car prices, we can develop better pricing strategies and production plans for the automotive industry.

Overall, the scope of car price prediction is to provide valuable insights and information to stakeholders in the automotive industry. By developing accurate prediction models, comparing algorithms, and identifying the most important features that influence car prices, we can help drive innovation, improve pricing strategies, and facilitate more efficient and transparent transactions in the automotive market.

Chapter 2:

Literature Review:

YEAR	AUTHOR	TITLE	OUTCOMES	DRAWBACKS
2019	Mustapha Hankar,Marouane Birjali	"Used Car Price Prediction using Machine Learning: A Case Study"	"Gradient boosting regressor showed a high R-squared score and low root mean square error"	Limited to a small dataset of only 200 samples, making it difficult to generalize the results
2022	B V Raghurami Reddy, Dr. K. Santhi Sree	"Car Price Prediction Using Machine Learning Algorithms"	Achieved 0.925595 R2 Score in predicting car prices	Used a relatively small dataset of only 1,000 samples, making it difficult to generalize the results.
2022	Rupali Satpute, Shreya Bhattacharjee	"Car Price Prediction Using Machine Learning Algorithms "	Found that the random forest algorithm performed the best, with an accuracy of over 95%.	Did not consider the impact of external factors such as economic conditions and gas prices on car prices.
2021	Prof. Borna Abramovic, Prof. Elżbieta Macioszek	"Price Prediction and Classification of Used-Vehicles Using Supervised Machine Learning"	Achieved an accuracy of 95% in predicting car prices using a random forest algorithm	Only few algorithms are used.

Overall, these studies demonstrate that the random forest algorithm can be effective in predicting car prices, but the accuracy of the predictions depends on the quality and quantity of data available.

Chapter 3:

Proposed System:

The proposed system for car price prediction will leverage the power of machine learning algorithms to provide an accurate and efficient method for predicting car prices based on various input features. The system will be designed to take in a dataset of historical car prices and features, preprocess the data using feature engineering techniques such as data normalization, categorical encoding, and feature selection, and train a machine learning model using the random forest algorithm. Once the model is trained, it can be used to predict the prices of new cars based on their features, and users can also input their own car data to receive a predicted price.

The proposed system will also provide a user-friendly interface for users to interact with the system, input car data, and view the predicted prices. The system will include features such as dataset management, model training and testing, prediction, visualization, feature engineering, customization, and integration with other software systems and tools. Overall, the proposed system for car price prediction will enable car dealerships, buyers, and sellers to make more informed decisions about car prices, and provide a valuable tool for the automotive industry.

3.1 Features and Functionality:

This system would provide a range of features and functionality to enable accurate and efficient prediction of car prices based on various input features, and would be a valuable tool for the automotive industry.

1. Input form: An input form with fields for users to input car features such as make, model, year, mileage, transmission type, fuel type, etc.
2. Prediction result: A display of the predicted price based on the input car features.
3. User management: Ability to manage user accounts, such as registration, login, and password reset.
4. User can predict the price based on various attributes of the car.

Chapter 4:

Requirement Analysis:

Functional requirements:

- The system should be able to predict the price of a car based on its features such as make, model, year, mileage, transmission type, fuel type, etc.
- The system should be able to visualize the dataset and model performance using techniques such as scatter plots, histograms, and learning curves.
- The system should be able to manage the dataset of car prices and features, including cleaning, preprocessing, and storing it in a database.
- The system should be able to train and test machine learning models on the dataset using techniques such as cross-validation and hyperparameter tuning.
- The system should be able to perform feature engineering tasks such as data normalization, categorical encoding, and feature selection to improve the accuracy of the predictions.
- The system should have a user-friendly interface for users to interact with the system, input car data, and view the predicted prices.

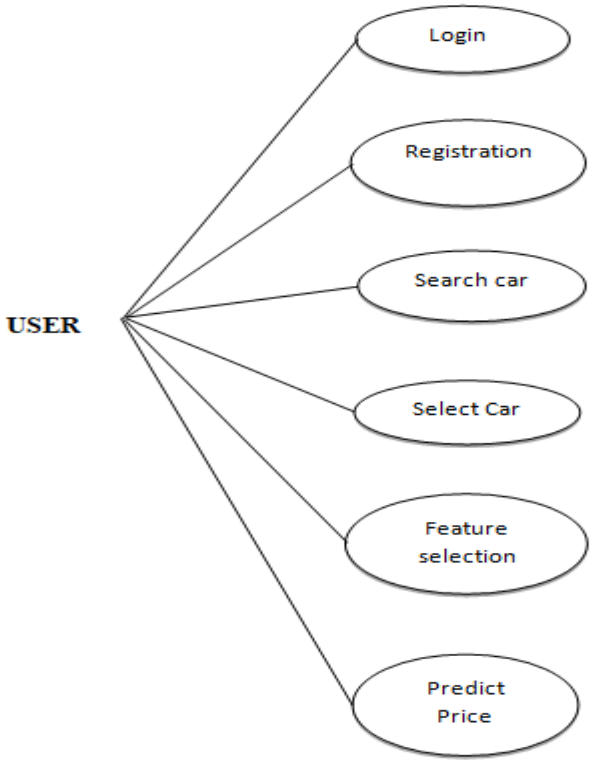
Non-functional requirements:

- The system should be able to handle large datasets and high traffic.
- The system should be fast and responsive, with minimal latency.
- The system should be scalable, able to grow or shrink as demand fluctuates.
- The system should be secure, with measures such as data encryption, user authentication, and access control.
- The system should be reliable, with minimal downtime and high availability.
- The system should be easy to maintain and upgrade, with clear documentation and modular architecture.
- The system should be compatible with various operating systems and hardware platforms.

Chapter 5:

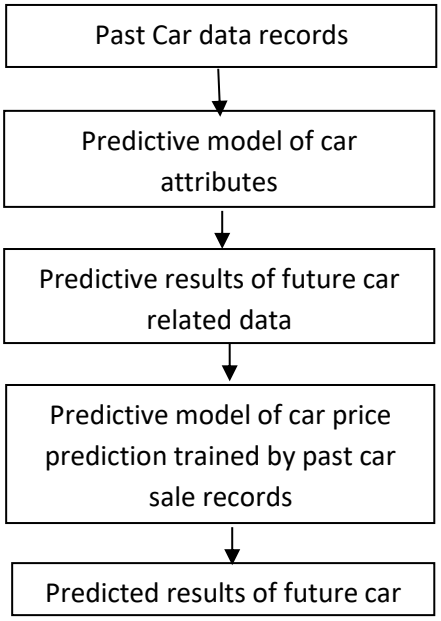
Project Design

5.1 Use Case Diagram



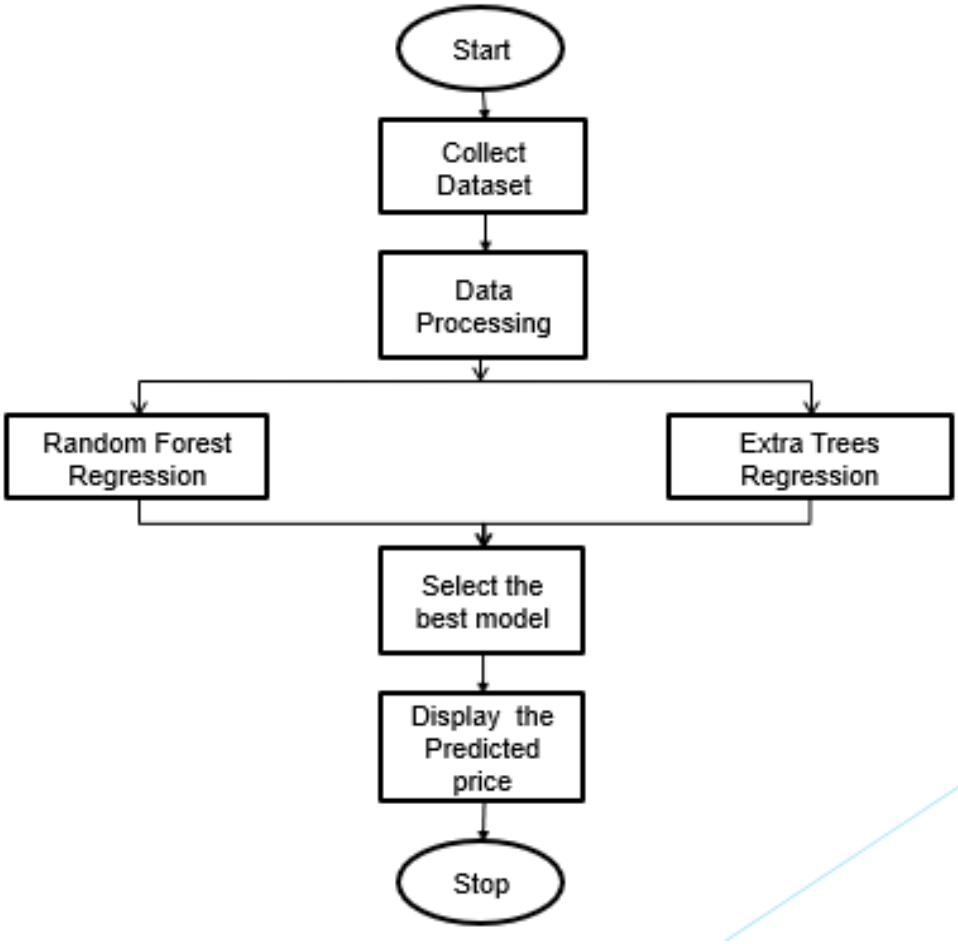
(Figure 5.1)

5.2 DFD



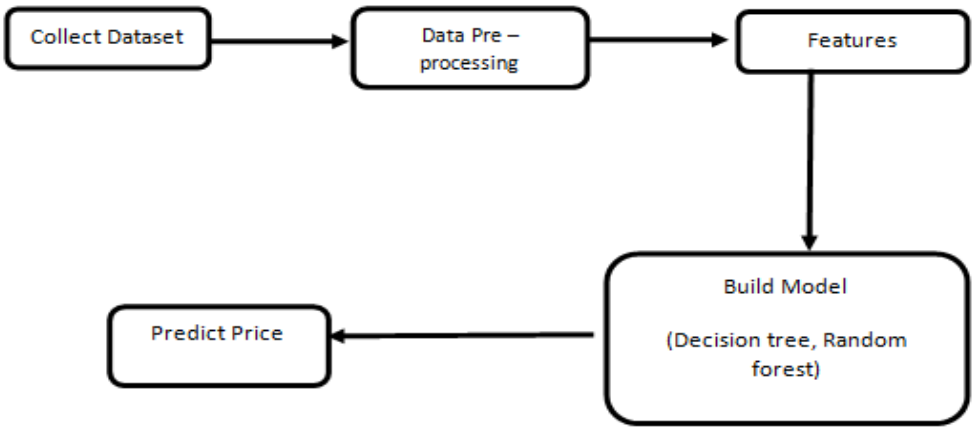
(Figure 5.2)

5.3 Flowchart:



(Figure 5.3)

5.4 System Architecture:



(Figure 5.4)

Chapter 6:

Technical Specification:

Frontend: Html, CSS, Python

As a web developer, the three main languages we use to build websites are HTML, CSS, and Python. Python is the programming language, we use HTML to structure the site, and we use CSS to design and layout the web page. These days, CSS has become more than just a design language, though. You can implement animations and smooth transitions with just CSS.

OS: Windows

Windows is a graphical operating system developed by Microsoft. It allows users to user to view and store files, run the software, play games, watch videos, and provides a way to connect to the internet. It was released for both home computing and professional works.

Backend: Flask, Mysql

Flask:

Flask is a web application framework written in Python. Armin Ronacher, who leads an international group of Python enthusiasts named Pocco, develops it. Flask is based on Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco projects.

MySQL:

MySQL is a very popular open-source relational database management system (RDBMS)MySQL is a relational database management system based on the Structured Query Language, which is the popular language for accessing and managing the records in the database. MySQL is a relational database management system based on the Structured Query Language, which is the popular language for accessing and managing the records in the database.

Development Environment: Jupyter Notebook

Jupyter Notebook:

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

Software Requirements:

Frontend:

• CSS
• HTML
• Python
• BOOTSTRAP

Backend:

- Flask
- MySQL

Development Environment: Jupyter Notebook

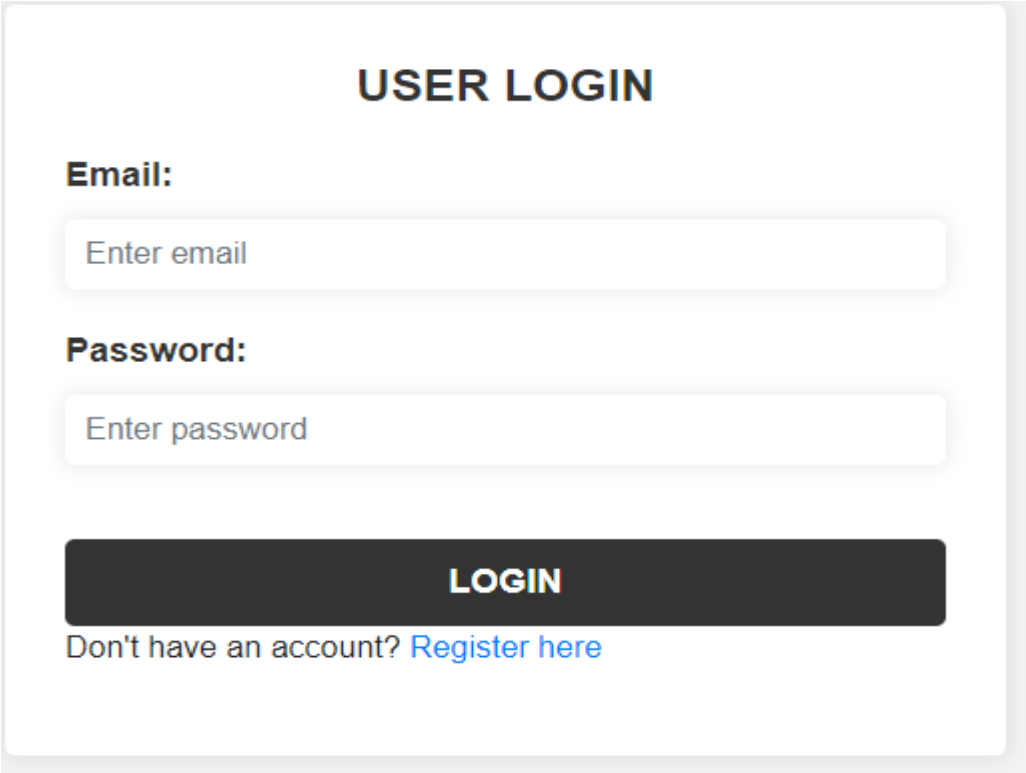
Chapter 7

Project Scheduling Template:

Sr. No	Group Member	Time duration	Work to be done
1	Shivam Gupta	1st week of january	Login page with database connection
	Yashab Mahimi		Testing login and registration page to find possible bugs
	Meet Bohra		Validation of login page.
	Kaif Khan		Registeration page
2	Yashab Mahimi	2nd week of february	Dataset Preparation and cleaning
	Meet Bohra		Implementing Home page
	Kaif Khan		Dataset processing
	Shivam Gupta		Bootstrap and testing of bugs
3	Meet Bohra	By the end of march month	Implementing result page
	Kaif Khan		Finding bugs and testing
	Yashab Mahimi		Flask routing and coding
	Shivam Gupta		CSS and data fitting

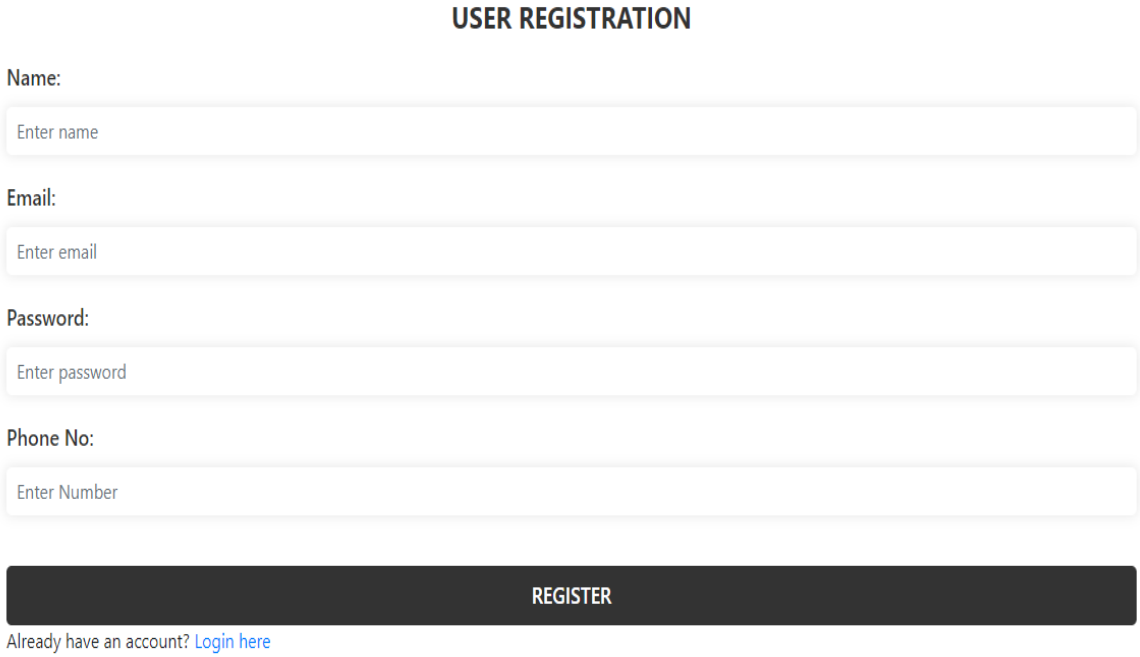
Chapter 8

Implementation:



(Figure 8.1)

Sign in page : User will login with his credentials on this page as shown in fig. 8.1.

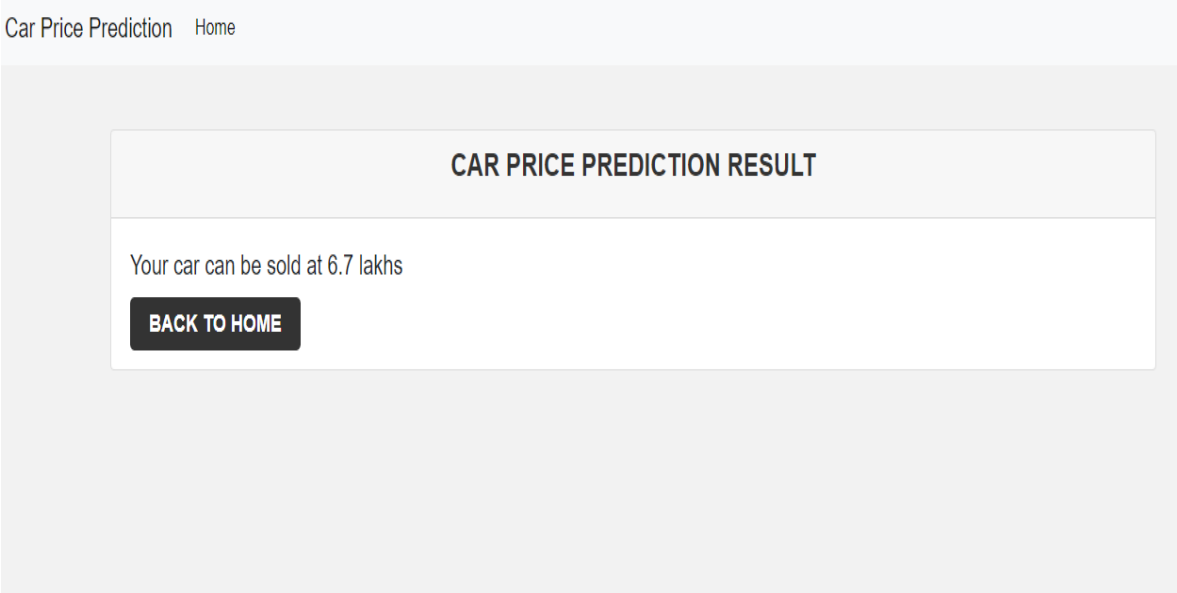


(Figure 8.2)

Sign up page: User can create his account on this page as shown in fig. 8.2.

(Figure 8.3)

Homepage: List of attributes will be displayed on this page. User can enter values of attributes of his car. as shown in figure 8.3.



(Figure 8.4)

Description page: Description page of each product is available. User can see product details using this page as shown in fig. 8.4.

Chapter 9

Results & Challenges

The application can be used for any Car prediction. It is easy to use, since it uses the GUI provided in the user dialog. User friendly screens are provided. The application is easy to use and interactive making online shopping a recreational activity for users. It has been thoroughly tested and implemented.

Challenges

- Compatibility with browsers like Mozilla Firefox, Internet explorer, Google Chrome etc.
- Using a layered approach in developing the application which would make the application maintainable.

The overall idea of doing this project is to get a real time experience. Learn modern technologies.

Chapter 10

Conclusion and Future Scope

Conclusion:

Car Price Prediction can be a challenging task due to high number of attributes that should be considered for accurate prediction. Our Prediction will help the customers to buy or sell their cars according to the market value.

Future Scope:

In future this machine learning model may bind with various website which can provide real time data for price prediction. Also we may add large historical data for car price which can help to improve accuracy of the machine learning model. 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.

References

[1] M. Hankar, M. Birjali and A. Beni-Hssane, "Used Car Price Prediction using Machine Learning: A Case Study," 2022 11th International Symposium on Signal, Image, Video and Communications (ISIVC), El Jadida, Morocco, 2022, pp. 1-4, doi: 10.1109/ISIVC54825.2022.9800719.

[2] https://www.ijraset.com/research-paper/car-price-prediction-using-machine-learning-algorithms.

[3] MDPI and ACS Style Bukvić, L.; Pašagić Škrinjar, J.; Fratrović, T.; Abramović, B. Price Prediction and Classification of Used-Vehicles Using Supervised Machine Learning. *Sustainability* 2022, *14*, 17034. https://doi.org/10.3390/su142417034

[4] Thilakarathne, Haritha. (2016). Developing a Hybrid Model For Disaster Prediction using Machine Learning with Artificial Neural Networks & Data Mining Approach. 10.13140/RG.2.2.11487.71845.