

**NORTH SOUTH UNIVERSITY**

**Department of Electrical & Computer Engineering**

**Project Report Spring 2023**

**Project Name: A multivendor e-commerce system of car price predictor using machine learning**

**Course ID: CSE498R**

**Course Title: Direct Research**

**Supervised under: Dr Dihan Md Nuruddin Hasan**

**Group Members:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Id** | **Section** |
| **Arafat Hossen** | **1831145642** | **15** |
| **Tamim Rahman** | **1831177042** | **24** |
| **Umme Hafsa** | **1911847642** | **24** |
| **Nusrat Jahan** | **1911657042** | **24** |

**Submission Date: 20th June 2023**

**Table Of Contents**

**1. Introduction…………………………....…………………………………………………….2**

**2. Problem Statement ………………………………………………………………………….2**

**3. Proposed Solution ……………………………………………………………………….......3**

**4. Related Work …………………………………………………………………………..........3**

**5. Technical Approach …………………………………………………………………...........5**

**6. Technology Platform …………………………………………………..………….……….11**

**7. Timeline ………………………………………………………………………….…...……12**

**8. Description of the Project output and discussion …………………….……………........13**

**9. Lesson learnt …………………………………………………………….……….….…….25**

**10. Contribution of individual team member ………………………….……..…………...25**

**11. Future Work……...…………………………………………………………..…………26**

**12. Conclusions …………………………………………………………………………......26**

**13. Reference …………………………………………………………….……………..…...27**

**1. Introduction**

Car Bazaaris basically a multivendor e-commerce website. A multi-vendor marketplace is a website, which hosts multiple vendors on its website and provides an opportunity for various vendors to sell online on a single platform. The owner of the website only manages the website and the third-party vendors get to sell their products online.

Basically, on this website, there are three types of actors. They are admin, seller, and buyer. Admin manages the website and invites more sellers. Firstly, Buyer creates their own account and after that, they can buy a car from this website. The seller also creates an account at first. Then they also give a post to sell their car. At this moment, it looks like a typical multivendor website but it has some extra features. We create a car price predictor system.

If anyone wants to sell their old car. They can take help from this website. In this project, the user gives the company name, model, purchase year, fuels, and number of kilometers that the car has traveled. Then it predicts the best possible price at which the user can sell the car. We take the dataset from Kaggle.

Basically, in this project, we will build a car price predictor using Linear Regression. We will also convert it into a full-fledged website using the flask framework.

The purpose of this project is to develop a Car Sales and price prediction System that allows customers to add their own vehicle information and also can purchase forms here. The customer must first watch the Car Sales System program in order to do so. it is difficult to maintain the car information individually and to supply for the customers who are eager to buy them. The customer has to face difficulty in order to know the information of the car, like manufacturing year, car model, and other valuable information in a single domain. Our main idea is to develop a system where we can have all the required information for the user in order to effectively interest him in the process of buying a car. Customers can rapidly obtain information such as car details that have been entered clearly using this application.

**2. Problem Statement**

Almost everyone wants their own car these days, but because of factors like affordability or economic conditions, many prefer to opt for pre-owned cars. Accurately predicting used car prices requires expert knowledge due to the nature of their dependence on a variety of factors and features. Used car prices are not constant in the market, both buyers and sellers need an intelligent system that will allow them to predict the correct price efficiently. In this intelligent system, the most difficult problem is the collection of the dataset which contains all important elements like the manufacturing year of the car, its gas type, its condition, miles driven, horsepower, doors, number of times a car has been painted, customer reviews, the weight of the car, etc. It is clear that the price of the product is affected by many factors, but unfortunately, information about these features is not always readily available. But with our application, all this information will be available when a user going to buy a car.

**3. Proposed Solution**

* Provides search facilities based on various factors. Such as Inventory, Car Models, Sales, Car Owners
* It tracks all the information about Customers, Cars, Sales, etc.
* Manage the information of Customers.
* Shows the information and description of the Inventory, Car Models
* To increase the efficiency of managing the Inventory, and Customers.
* It deals with monitoring the information and transactions of Sales.

**4. Related Work**

Every day, businesses need to decide how to allocate valuable resources based on predictions. Unfortunately, whilst most practitioners recognize the importance of accurate predictions of development effort for tendering bids, monitoring progress, scheduling resources, and evaluating risk factors, current estimation techniques are often highly inaccurate. In this article, we compared three machine learning techniques with an LSR model to predict the efforts of a software project. These techniques were compared in terms of accuracy, explanatory values, and configurable. We have found differences in the accuracy of the predictions, but argue that these methods can have an equal impact on adoption. You can see that the explanatory values ​​for both estimates are by analogy [1].

Electric vehicles (EVs) are considered a promising alternative to traditional fossil fuel vehicles to Reduce CO2 emissions and pollution. but, Global interest and investment, user recruitment It's still low. Above all, range anxiety applies as One of the main barriers the to large-scale adoption of electric vehicles. This is defined as the possibility of being weakened and present It is turned off in the middle of the trip. The hybrid model is Excellent performance in processing high-dimensional data Internet of Things (IoT), especially for reaching Estimates of EVs, which has recently been regarded as an important IoT platform. The main goal of SOM is Converts input signal patterns of any dimension to Convert to a 2D discrete map and transform adaptively in a topologically ordered manner. However, traditional feature selection is not included SOM and knowledge extraction Trained codebook, and stained neurons. on the other hand, Regression tree (RT) models can provide automated functionality Selection by evaluating the information obtained in the training set. But when applied globally, bushy trees are created and lost Interoperability. A hybrid model including the above Two can easily make up for other weaknesses and Provides range-optimized and economical solutions for anxiety disorder [2]

Understanding With travel demand Population car ownership is a central concern of city officials from all over the world. Transportation and urban planner Notify policies using traffic demand forecasts an investment that puts a high value on data and methods that Provides demand forecast. Discrete choice model is a set of econometric tools Widely used to describe transport behavior. Including the household's decision to own a car. The general goal of this type of modeling is to be unbiased. An estimator that provides behavioral and economic insights. An Accurate discrete choice model fits observational data with Minimal error. Normally, we assume a linear model structure. A discrete choice model is selected for bidding Interpretability.

The discrete choice model describes the decisions made by the decision-maker. Examine the situation where the possible results are discrete among the choices. Decision-makers can be people with households, businesses, or other decision-making bodies the alternative could be a course from a competing product, Actions, or other options or items for which a decision must be made. The result is machine learning and econometrics Need a model with different purposes and mechanisms A dataset that has been preprocessed differently. Not pre-treating or do not perform feature engineering to get a typical dataset It’s a machine learning method, but it’s training the machine. Train the model with discrete choice model data. It doesn't take full advantage of machine learning models. The results show that demand forecasting is needed in the context of transportation, machine learning methods can contribute to this. We can support cities and transportation planners [3].

A very interesting area of ​​research that has stood out in recent years is the detection and prediction of objects based on features that can benefit consumers and the industry. In this article, you will understand the concepts of object detection, such as car detection, to find out the price of a used car using automated machine learning techniques. You also understand the concept of object detection categories. The most difficult task today is to determine the list price of used cars on the market. Possible factors that can affect the price of a used car. The main goal of this task is to develop a machine learning model that can accurately predict the price of a used car based on parameters and characteristics. This document uses an implementation and valuation method for an automotive dataset consisting of the selling prices of different models in different cities in India. The results of this experiment show that linear regression and clustering using a random forest model give the highest accuracy results. Machine learning models produce satisfactory results in a short amount of time compared to those self-described above [4].

Urban populations are growing rapidly due to rising urban living standards. Smart 24 City consists of various submodules such as smart home, 25 smart waste management, smart power supply, 26 smart water management, smart transportation, and more. Due to improved living standards, the rural population of 28 is increasingly moving to urban areas. The number of vehicles in smart cities is increasing rapidly. This paper proposes a secure IoT-based architecture for intelligent vehicle communications. Use surveillance machine learning algorithms to test and validate models on three 120 road accident datasets. With the development of information and communication technology (ICT) and other new technologies such as the Internet of Things (IoT), cloud computing, and artificial intelligence (AI), the transportation system is now called "intelligent" transportation. increase. The number of vehicles has also increased rapidly causing many traffic jams and accidents on a daily basis. Therefore, predicting the severity of a traffic accident is very helpful in developing intelligent transportation systems. It's time to implement this on the road to save lives. Various machine learning models are used in this work, and it is concluded that the decision tree model has the highest average accuracy [5].

Car Bazaar is basically a multi-vendor e-commerce site. A multi-vendor marketplace is a website that hosts multiple vendors on its website and offers different vendors the opportunity to sell online on a single platform. Website owners control only the website, and third-party vendors can sell their products online. There are basically three types of actors on this website. You are an administrator, a seller, and a buyer. The administrator manages the site and invites more sellers. First, buyers can create their own accounts and then buy a car on this website. The seller also creates an account first. Then they also give mail to sell their car. At the moment it looks like a typical multi-vendor website, but with some additional features. Create a car price forecasting system. If someone wants to sell an old car. You can get help from this website. In this project, users will provide the company name, model, year of purchase, fuel, and the number of kilometers the car has traveled. Then predict the best price the user can sell the car. Use Kaggle's dataset. Basically, this project uses linear regression to create a car price predictor. It also uses the Flask framework to convert to a full-fledged website. From the above analysis, we can infer that the main difference between our projects is that our project has both buy and sell options with predictive options. Finally, we hope that people can use our application to choose the best car for them.

**5. Technical Approach**

Aims to create a strong machine learning model that provides precise price prediction of a car using linear regression. To estimate the price of an automobile based on factors such as price, year of purchase, mileage, model, etc., we needed a previous data collection of previously owned vehicles. These characteristics are regarded as independent variables, although the cost of the car depends on them. Our attributes are combined in the equation of linear regression, which then reveals the outcome.

Y = β0 + β1X

In the equation 0 above, Y is the output or necessary variable, X is the input, and 1 denotes the regression coefficients. The equation above represents the relationship in the case of a single input. The equation provided below is used when the user provides numerous inputs.

Y = β0 + β1X1+ β2X2+……. βnXn  
There are several characteristics and aspects that affect a car's price, which ultimately produces large data sets that are difficult to analyze. Our study is focused on creating a model that can handle highly complex algorithms and produce reliable results regardless of the size of the data collection. The input data set, which was gathered over the course of three to four months from the internet marketplace Quikr, includes both luxury vehicles and everyday commuter vehicles. Variable inputs such as color, rims, body type, transmission, engine size, register city, power steering, alloys, mileage, etc. were included in the data collection, which is extremely unclean data. After the data gathering was finished, we cleansed the data, which provided the qualities that we used to build our system. To anticipate the cost of a given car, we analyzed the data using a multivariate linear regression technique.

To reduce the time required for training, we used 892 thousand examples from our dataset. Linear Regression, Random Forest and OneHotEncoder [6].

**1****. Lenear Regression:**

Due to its simplicity and relatively quick training time, linear regression was chosen as the first model. The features were used straight as the feature vectors without any feature mapping. Because of the results' obvious low variance, regularization was not used [7].

**2. Random Forest:**

An ensemble learning-based regression model is Random Forest. In order to create the ensemble model, which collectively yields a prediction, it employs a decision tree model, specifically as the name implies, several decision trees. The advantage of this model is that the trees are created simultaneously and with little correlation, which leads to good outcomes because no one tree is susceptible to individual errors of other trees. The use of Bootstrap Aggregation or bagging, which provides the randomness necessary to create resilient and uncorrelated trees, helps to partially ensure this uncorrelated behavior. Therefore, this model was chosen to compare a bagging strategy with the following gradient boosting methods and account for the dataset's huge number of features.

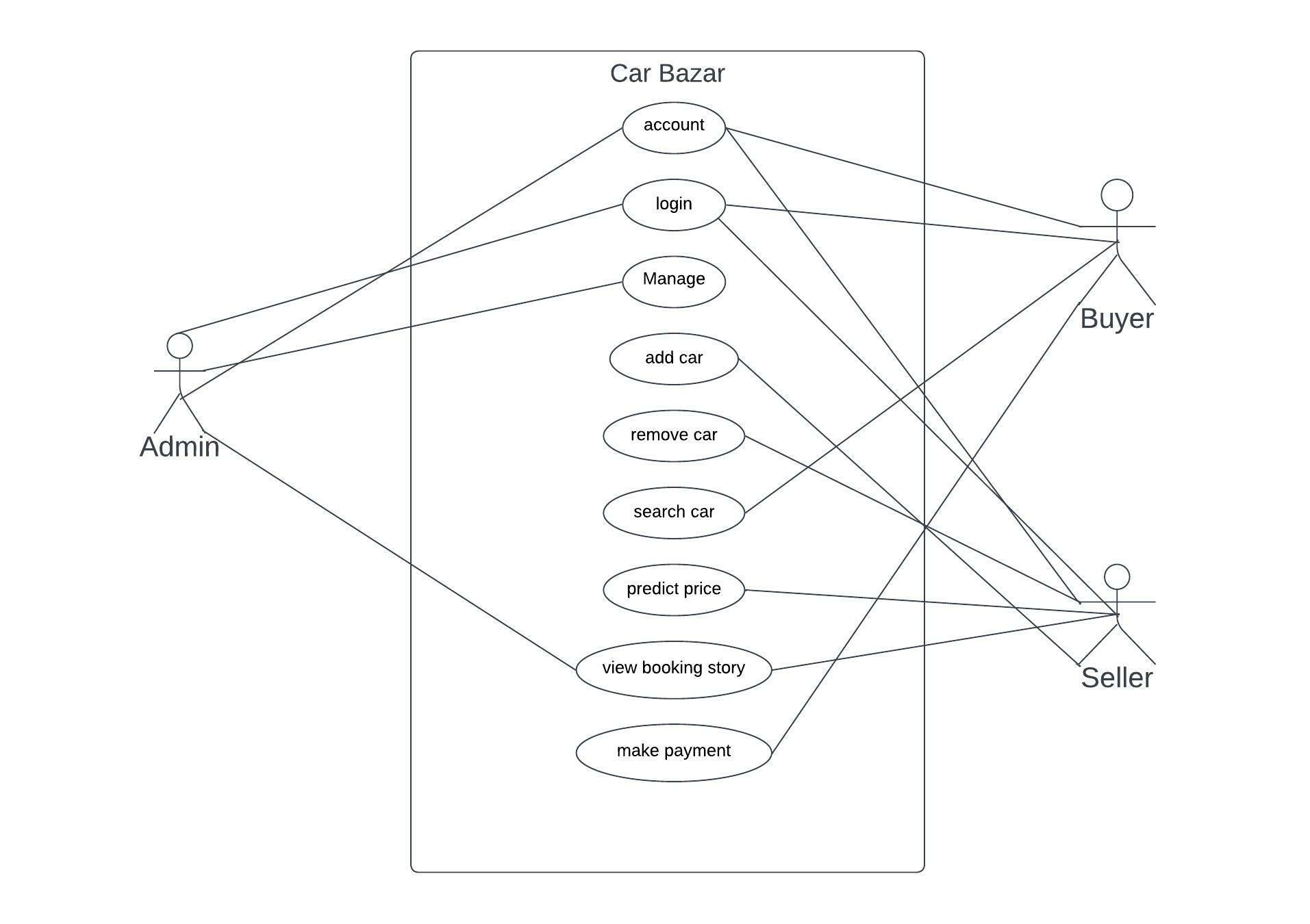
**3****. OneHotEncoder.**

Machine learning algorithms cannot work with categorical data directly. Categorical data must be converted to numbers. This applies when you are working with a sequence classification type problem and plan on using deep learning methods such as Long Short-Term Memory recurrent neural networks. One hot encoding is a representation of categorical variables as binary vectors. This first requires that the categorical values be mapped to integer values. Then, each integer value is represented as a binary vector that is all zero values except the index of the integer [8].

**4.** **R2 Score:**

The R2 score is one of the performance evaluations measures for regression-based machine learning models. It is also known as the coefficient of determination. If you want to learn how to evaluate the performance of a machine learning model using the r squared score. It is pronounced as R squared and is also known as the coefficient of determination. It works by measuring the amount of variance in the predictions explained by the dataset. Simply put, it is the difference between the samples in the dataset and the predictions made by the model. It is never used the r squared score when evaluating the performance of a regression-based machine learning model, it can learn more about its implementation using Python [9].

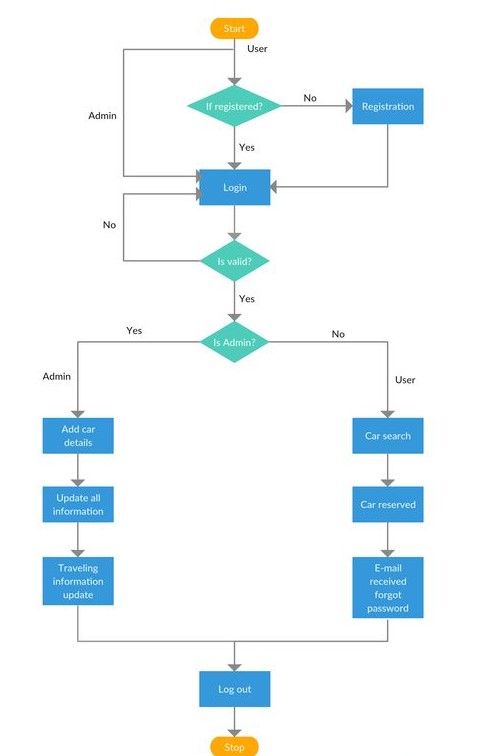
**Design Implementation**

**Tentative schematic diagram**

**Figure 1: Use case Diagram**

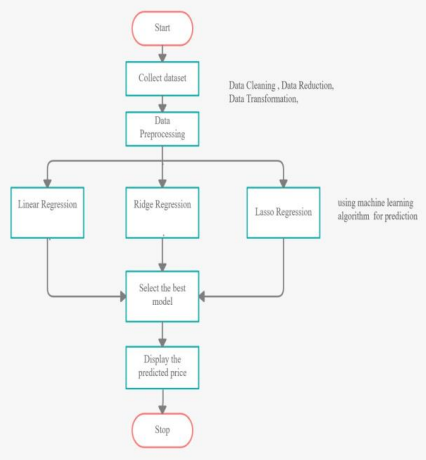
Here on this diagram there are two part admin who manage all the administration task on the other hand user can got the access when they regrester with this application. In tentative diagram there are three parts Admin Buyer and Seller. Here, Admin can acess and manage all the section. For the Buyer acess buyer must need create a accout.By this account Buyer can acess acount,login, search car and make payment. Seller can aslo need a account, by this account they can aslo acess some sention such as acount, login, add car, remove car, predict price and view boking story.

**Flowchart:**

****

**Figure 3: Flowchart 1**

This our car buys sell flowchart diagram. There is the step by step process. Using this diagram helps to create a car booking service system for the customers.

****

**Figure 4: Flowchart 2**

As shown in the above figure, the process starts by collecting the dataset. The next step is to do Data Preprocessing which includes Data cleaning, Data reduction, Data Transformation. Then, using various machine learning algorithms we will predict the price.

**6. Technology Platform**

1. **PyCharm:** PyCharm is an [integrated development environment](https://en.wikipedia.org/wiki/Integrated_development_environment) (IDE) used in [computer programming](https://en.wikipedia.org/wiki/Computer_programming), specifically for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) programming language. We use python language for this. That’s why we use PyCharm IDE.

2. **HTML, CSS, and Bootstrap**: We use HTML, CSS, and bootstrap for our

project’s frontend part. The **Hyper Text Markup Language** or **HTML**

is the standard [markup language](https://en.wikipedia.org/wiki/Markup_language) for documents designed to be displayed

in a [web browser](https://en.wikipedia.org/wiki/Web_browser). It can be assisted by technologies such as (CSS) and

[scripting languages](https://en.wikipedia.org/wiki/Scripting_language) such as [JavaScript](https://en.wikipedia.org/wiki/JavaScript).

3. **Python:** Python is a high-level, interpreted, general-purpose programming

language. Its design philosophy emphasizes code readability with the use of

significant indentation. Python is dynamically-typed and garbage-collected.

It supports multiple programming paradigms, including structured (particularly

procedural), object-oriented and functional programming. In this project we use

Python.

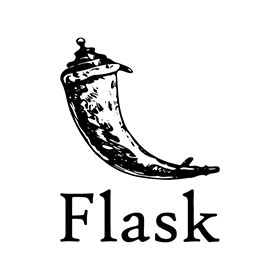


4. **Django:** Django is a free and open-source, Python-based web framework that

follows the model–template–views (MTV) architectural pattern. Django's primary

goal is to ease the creation of complex, database-driven websites. For this, e-commerce

the part we use Django framework.

****

**5. Flask:** Flask is a micro web framework written in Python. It is classified as a

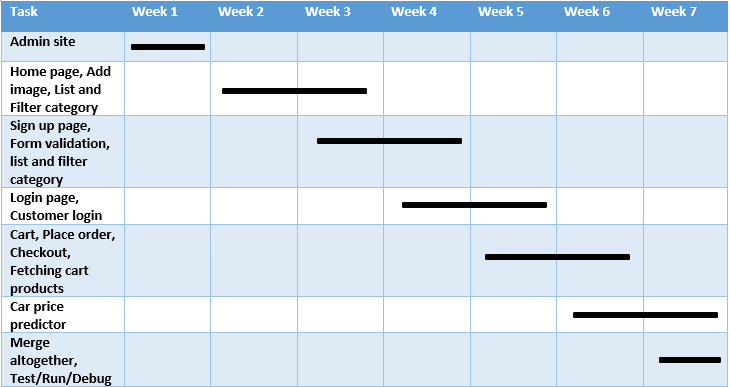
microframework because it does not require particular tools or libraries. It has no

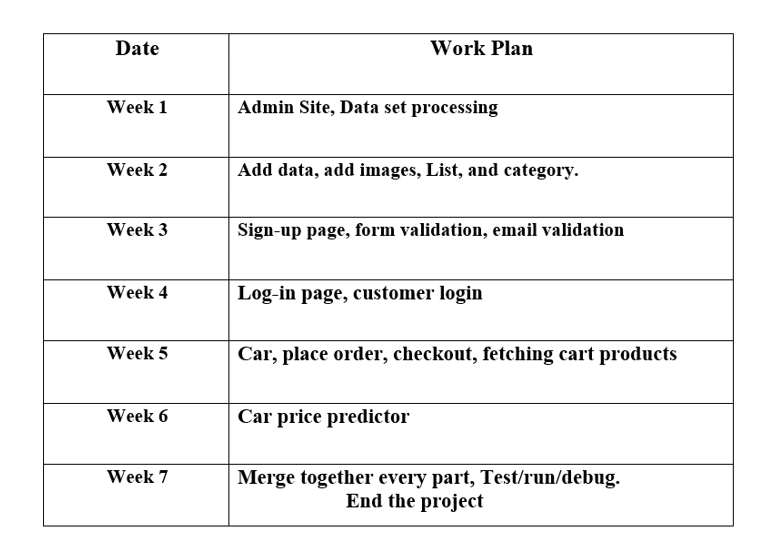
database abstraction layer, form validation, or any other components where pre-

existing third-party libraries provide common functions. In our project’s price

predictor part, we use this framework.

**7. Timeline**





**8. Description of the Project output and discussion**

Following machine learning classifiers are implemented.

1. Data collection

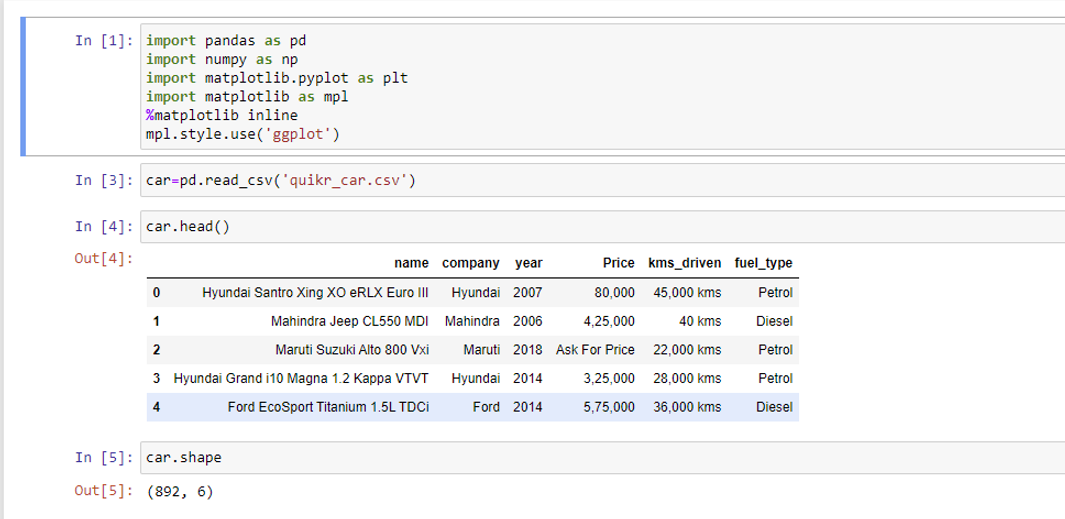
2. Data cleaning

3. Exploratory Data Analysis

4. Regression Model and Evaluation

**1. Data collection**

Data was collected and Scrapped from a website named KAGGLE DATASETS, through multiple runs and irritations 892 rows of data with 6 variables unsorted data collected successfully from the website. At first, the data types of each attribute were corrected/converted by performing pre-processing on each attribute individually. Details and description of dataset is described in table below:

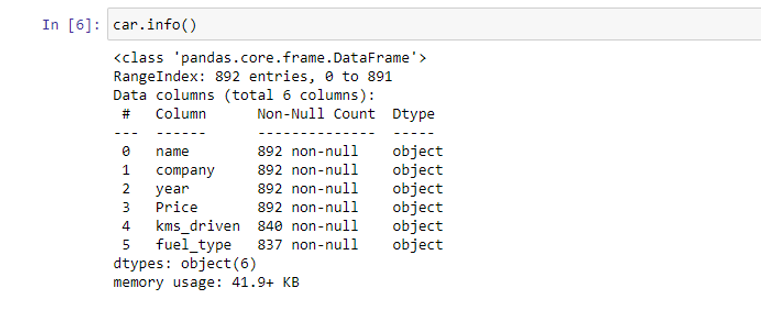


**Figure 5: Datasets**

**2. Data cleaning:**

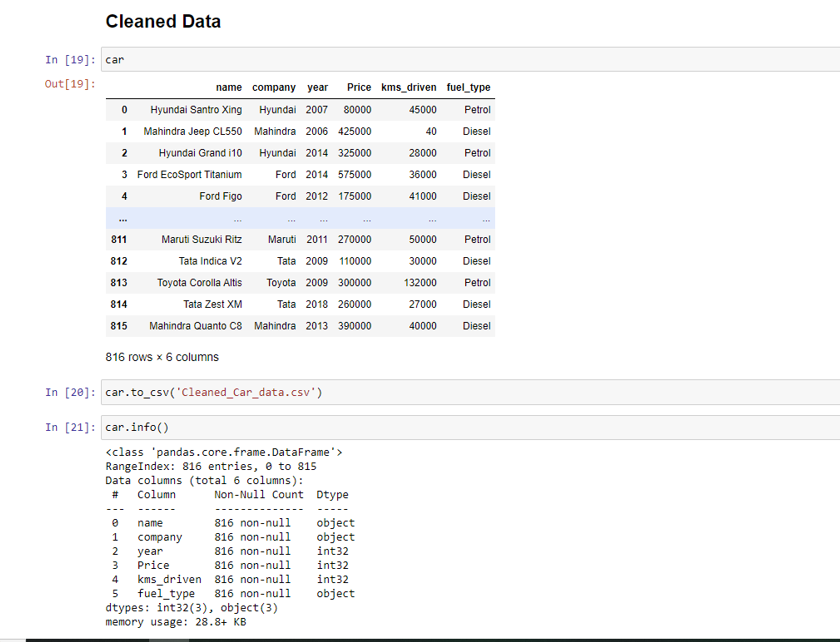
Scraped data collected from the website saved as csv file.

Here, need to look for missing values and check data types before trying to do any analysis:

Those data were unsorted,  


**Figure 6: Unsorted Data**

Then we sort the data with categories all the data and the sorted data saved as a csv file.

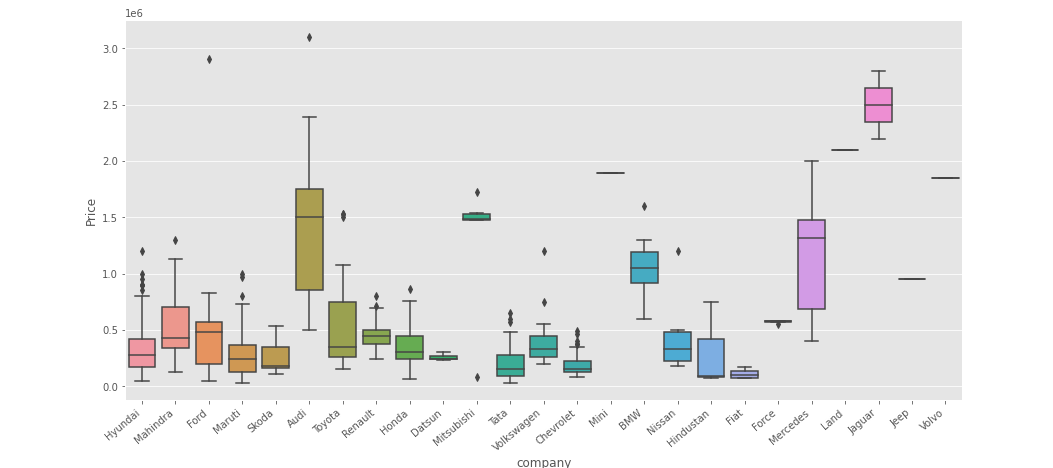


**Figure 7: Sorted Data**

**3. Exploratory Data Analysis (EDA):**

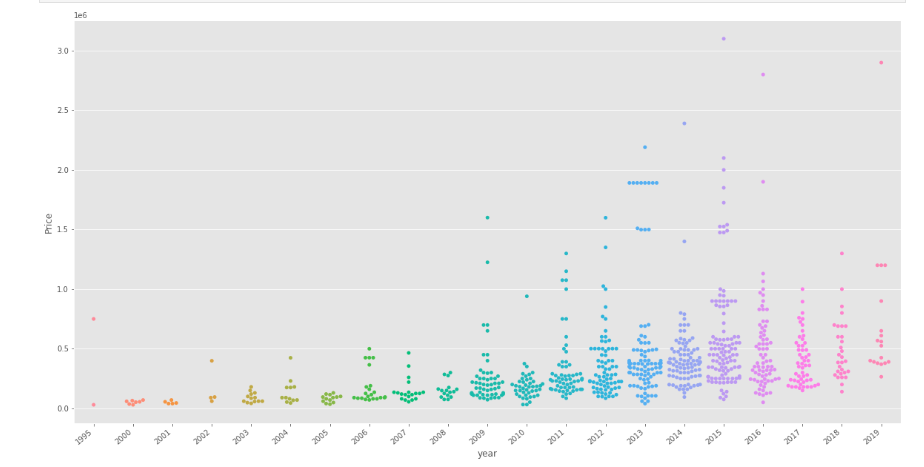
Due to its combination of Model and Year columns, Car Name, km/hrs., price was a redundant attribute, and it was removed. Exploratory Data Analysis composed of following steps:

**1.Checking relationship of Company with Price**



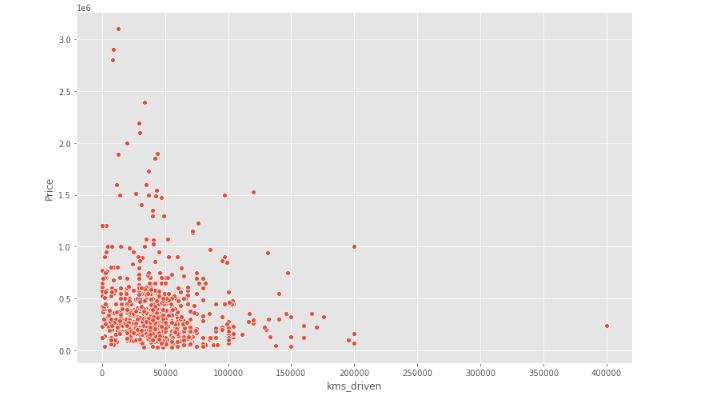
**Figure 8: Relationship of Company with Price**

**2. Checking relationship of Year with Price**



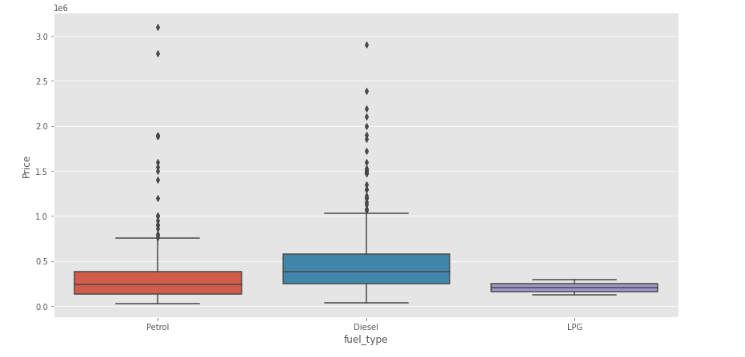
**Figure 9: Relationship of Year with Price**

**3. Checking relationship of kms\_driven with Price**



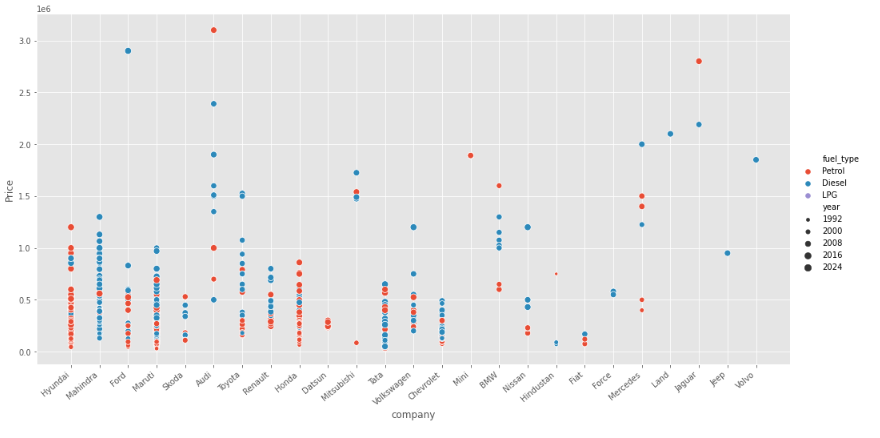
**Figure 10: Checking relationship of kms\_driven with Price**

**4. Checking relationship of Fuel Type with Price**



**Figure 11: Relationship of Fuel Type with Price**

**5. Relationship of Price with FuelType, Year and Company mixed**



**Figure 12: Relationship of Price with FuelType, Year and Company mixed**

**4. Regression Model and Evaluation**

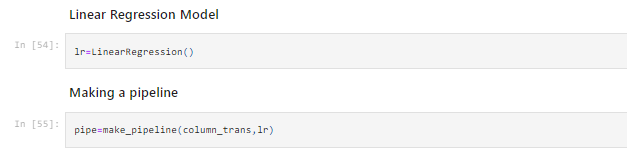
Linear regression is a widely-used supervised learning algorithm to predict a continuous dependent (or target) variable. Depending on the number of independent variables, it could be in the form of simple or multiple linear regression. Here created a multiple linear regression model because we used many independent variables to predict the dependent variable which is the price of a used car. After checking the correlation and distribution of variables, we decided to use Year, km, fuel to predict the price of a used car.

I used [scikit-learn](https://scikit-learn.org/stable/" \t "_blank) which provides simple and effective machine learning tools.



**Figure 13: Train Test Split**

Then we created a LinearRegression() object, trained it with train dataset.



**Figure 14 : Liner Regression Function**

Fitting the model by pipeline function



**Figure 15: Pipeline Function**

Checking R2 Score for Finding the model with a random state of TrainTestSplit where the model was found to give almost 0.92 as r2\_score *[10]*



**Figure 16: R2\_Score Function**

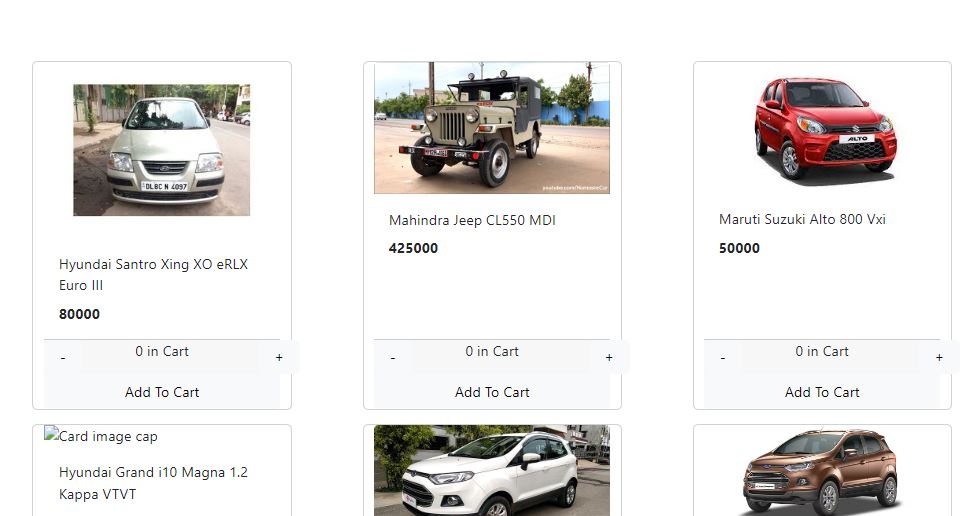
Then we found the best model is found at a certain random state



**Figure 17: Best Model**

Project interface description:

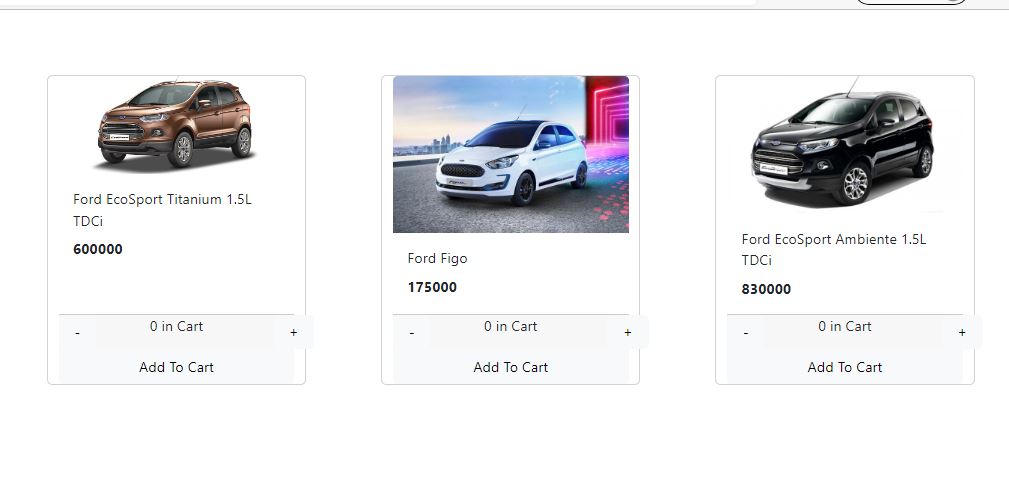
**1.Main Interface:**

****

**Figure 18: Main Interface**

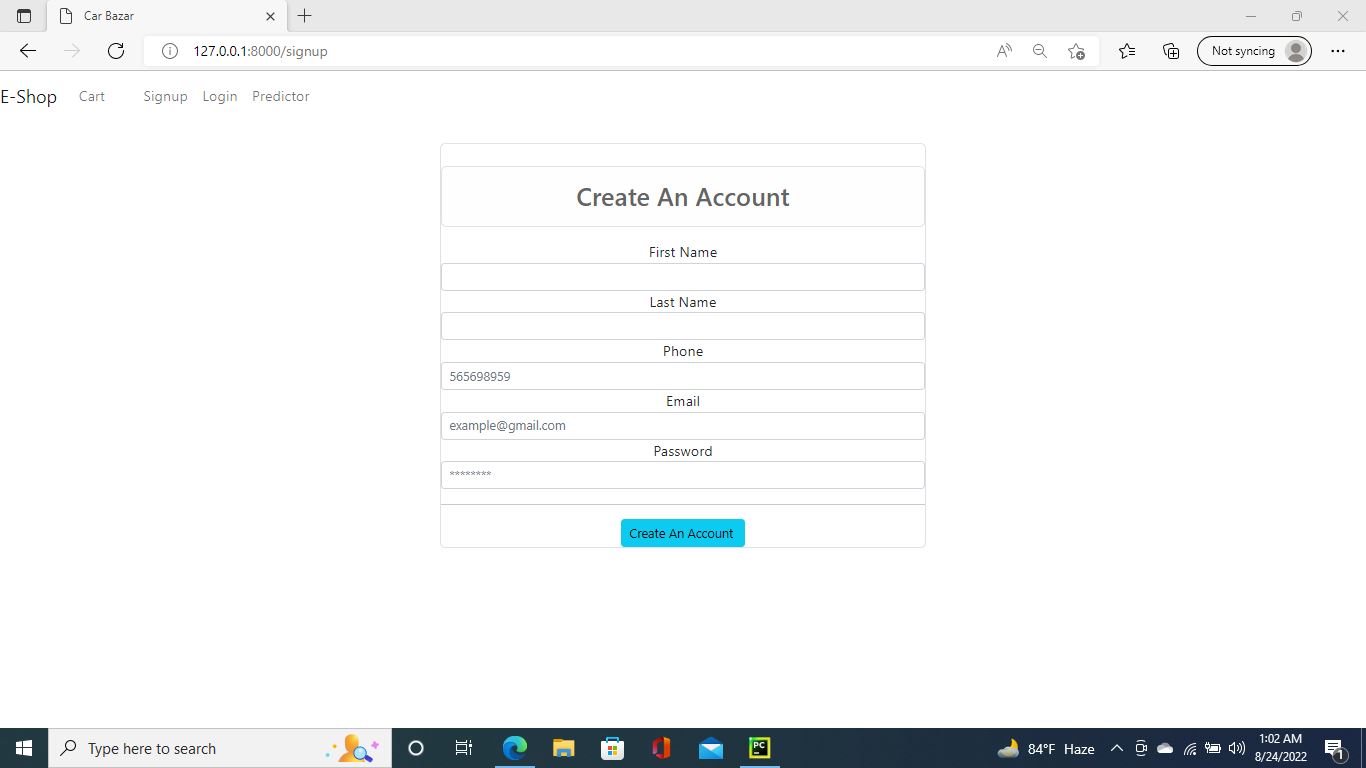
This our project user interface. List of all categorical cars.

**2.Interface after sorting:**

****

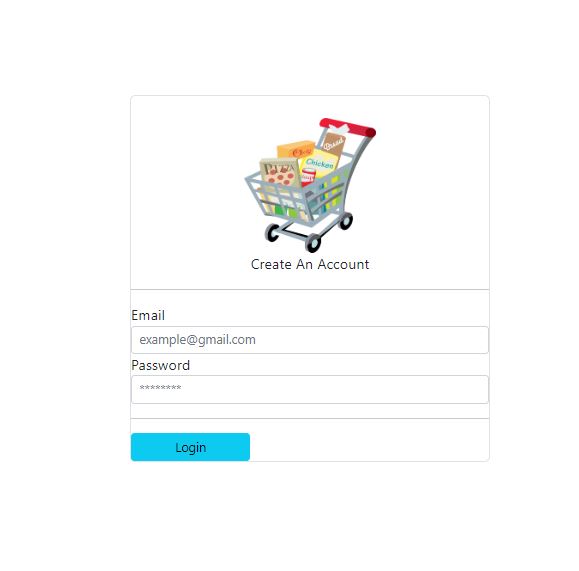
**Figure 19: Interface after sorting**

**3. Sign Up page:**

****

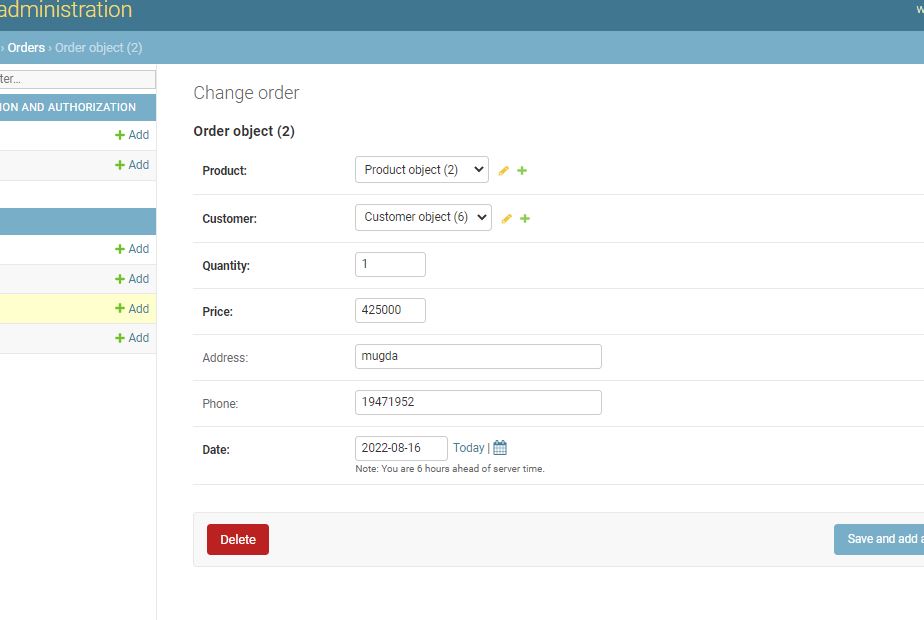
**Figure 20: Sign Up page**

**4.LogIn page:**

****

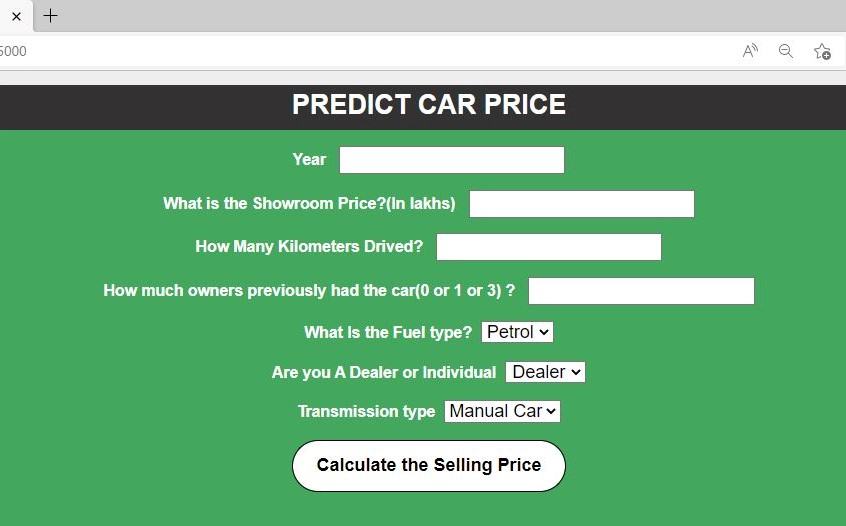
**Figure 21: LogIn page**

**5.Database:**

****

**Figure 22: Database**

**6. Predictor:**

****

**Figure 23: Predictor**

**9. Lesson learnt**

Challenges: While collecting the data here we face the data type problem which is numerical data and categorical data type. After collecting the data, we need to pre-process the data set. Here some

null and missing data are also available on the dataset. So the need to fix this also. Here we face a problem with the environment setup for the front end and backend part. Also, face some problems with the email validation part which is connected with the login purpose. Face a problem with adding the car to the cart list. Here we sort the car with the brand and configuration. Facing problems with this part also. While building those models like Linear regression, OneHotEncoder, and R2 Score. we face problems with the implementation. Because this is the first time we work

machine learning models. Here we also face a problem with the model accuracy. we face a problem with our machine-learning model. While building those models here faced problems with model accuracy which is the main part of the machine learning model.

**10. Contribution of individual team member**

The following shows the approximate contributions in the percentage of each member of our team while working on this project:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Requirement Analysis** | **Planning** | **System**  **Analysis** | **Design** | **Implementation** | **Testing &**  **Development** |
| **Tamim** | 30% | 30% | 30% | 30% | 30% | 30% |
| **Arafat** | 30% | 30% | 30% | 30% | 30% | 30% |
| **Hafsa** | 20% | 20% | 20% | 20% | 20% | 20% |
| **Nusrat** | 20% | 20% | 20% | 20% | 20% | 20% |
| **Total** | 100% | 100% | 100% | 100% | 100% | 100% |

**11. Future work**

Deep learning classifiers will be explored in the future when additional data is gathered using various web-scraping methods. We'll test algorithms like Quantile Regression, ANN, and SVM. The intelligent model will then be included in online and mobile applications for general use. Additionally, following the data collection phase, shortages of semiconductors have developed as a result of the pandemic, which caused a rise in car costs and had a significant impact on the used car market. the need for regular Data Periodically, data gathering and analysis are needed; ideally, we would have a real-time processing application.

**12. Conclusions**

With the increased prices of new cars and the financial incapability of customers to buy them, Used Car sales are on a global increase. Therefore, there is an urgent need for a Used Car Price Prediction system whicthatectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of a used car price prediction.   The Purpose of the application is to build an application program to reduce the manual work of managing the inventory of car shops, customers, car models, etc. Its tracks all the details about the car models, sales, and car owners. In this project, the user gives the company name, model, purchase year, fuels, and the number of kilometers that the car has traveled then it predicts the best possible price at which the user can sell the car.

**12. References**

|  |  |
| --- | --- |
| [1] | CarolynMair, "Journal of Systems and Software," Science Direct, September 2013. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0164121200000054. |
| [2] | IEEE, "A Hybrid Machine Learning Model," IEEE Global Communications Conference (GLOBECOM), December 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7841506. |
| [3] | E. H. Miguel Paredes, "Machine learning or discrete choice models," IEEE International Conference, June 2017. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8005618. |
| [4] | E. k. Anu Yadav, "Object detection and used car price predicting analysis system," LingCure , November 2021. [Online]. Available: http://www.lingcure.org/index.php/journal/article/view/1660. |
| [5] | B. K. Mohanta, "Machine learning based accident prediction in secure IoT enable transportation system," Digital transformation through advances in artificial intelligence and machine learning, January 2022. [Online]. Available: https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs189743. |
| [6] | S. Yıldırım, "Predicting Used Car Prices with Machine Learning," A complete data science project from data collection to model evaluation, January 2020. [Online]. Available: https://towardsdatascience.com/predicting-used-car-prices-with-machine-learning-fea53811b1ab. |
| [7] | "Areas of machine learning education," Tensorflow, [Online]. Available: https://www.tensorflow.org/resources/learn-ml?gclid=CjwKCAjwmJeYBhAwEiwAXlg0AZRjA36lIDHw5R\_Eo6hjmlr9bP7lGYX-UKsM4zkxBEgkfCCqm-wNChoCpG4QAvD\_BwE. |
| [8] | J. Brownlee, "One Hot Encode Sequence Data in Python," Machine Learning Mistry, August 2019. [Online]. Available: https://machinelearningmastery.com/how-to-one-hot-encode-sequence-data-in-python/. |
| [9] | A. KHARWAL, "R2 Score in Machine Learning," MACHINE LEARNING, June 2021. [Online]. Available: https://thecleverprogrammer.com/2021/06/22/r2-score-in-machine-learning/#:~:text=The%20R2%20score%20is%20a,predictions%20explained%20by%20the%20dataset.. |
| [10] | E. Gokce, "Predicting Used Car Prices with Machine Learning Techniques," Comparing Performance of Five Different ML Models, January 2020. [Online]. Available: https://towardsdatascience.com/predicting-used-car-prices-with-machine-learning-techniques-8a9d8313952. |
| [11] | "Journal of Systems and Software," September 2013. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0164121200000054. |
| [12] | "IEEE Global Communications Conference (GLOBECOM)," December 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7841506. |
| [13] | "International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)," June 2017. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8005618. |
| [14] | "Digital transformation through advances in artificial intelligence and machine learning," January 2022. [Online]. Available: https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs189743. |
| [15] | "Deep Learning, Machine Learning," [Online]. Available: http://www.lingcure.org/index.php/journal/article/view/1660. |