CONTENTS

DOMAIN SELECTION

The automotive valuation landscape is inherently dynamic, influenced by multiple factors. Machine learning, with its adaptability and capacity to recognize patterns within vast datasets, emerges as the ideal domain to address the intricacies of predicting car prices accurately.

The decision of selecting the domain depended on the following:-

1. DYNAMIC NATURE OF AUTOMOTIVE VALUATION : The dynamic nature of the automotive industry demands adaptive and agile solutions to navigate evolving market trends and consumer preferences.

2. UNLEASHING PREDICTIVE POWER : the Random Forest algorithm in our project, offers predictive power that goes beyond the limitations of conventional approaches.

3. ADAPTABILITY TO DIVERSE FACTORS : accommodates diverse factors, from traditional attributes to emerging trends, amplifying the model's predictive capability.

4. PRECISION & ACCURACY : Machine learning techniques excel in capturing complex patterns, resulting in a predictive model that attains unparalleled precision and accuracy in forecasting car prices

5. TRANSPARENCY AND FAIRNESS : The model learns and adapts from a diverse dataset, mitigating biases and contributing to a fairer pricing mechanism, addressing concerns of overcharging and promoting ethical practices.

6. SCALABILITY AND REAL-TIME ADAPTATION : Machine learning models inherently scale with the volume of data. As the dataset grows, our model's capacity to adapt and refine predictions in real-time ensures its relevance and accuracy.

7. USER-CENTRIC PREDICTIONS : Our application's intuitive interface, empowered by these domains, ensures users can effortlessly input diverse factors, making the car price prediction process accessible and user-friendly.

TITLE SELECTION & JUSTIFICATION

- The process of choosing a title for this project holds more significance than the label alone. It is a deliberate fusion of "Car" to signify the automotive landscape and "Quintessence" to convey the project's core values. This amalgamation reflects our project's holistic approach, encapsulating the diverse elements contributing to our predictive model, from advanced algorithms to seamless web integration.

- It serves as a beacon, symbolizing the project's mission of providing users with an essential and accurate tool for informed decision-making in car purchases.

KEY CONSIDERATIONS -

- MACHINE LEARNING : Using Machine learning algorithms, was driven by its ability to handle complex datasets and provide accurate predictions. Its versatility and capability to mitigate overfitting make it an ideal choice for our diverse car pricing.
- WEB INTEGRATION : The integration of this project with the web ensures a user-friendly and visually appealing platform. It collectively contributes to an optimal and engaging user experience. The technologies seamlessly integrate with each other, fostering a cohesive ecosystem. It also ensures real-time access to data, enabling efficient model training and prediction.
- SCALABILITY : Technologies are scalable, capable of accommodating a growing dataset and user base.

ABSTRACT

- In the dynamic landscape of India's automobile market, where transparency in car pricing is imperative, this project introduces a transformative approach using the Random Forest algorithm.

- The study harnesses a robust dataset comprising over 6000 elements, significantly surpassing previous endeavors. The larger dataset contributes to heightened predictive accuracy, marking a substantial advancement in comparison to existing projects.

- A pivotal aspect of our initiative lies in harnessing the power of machine learning models and the integration with web applications and the advancements in predictive modeling is at the core of our initiative.

- Web integration ensures an effortless and interactive user experience, allowing users to input diverse factors influencing car prices easily. This user-friendly interface enhances accessibility and provides clear and transparent price predictions, facilitating a rich user experience.

- This enables us to achieve unprecedented accuracy levels, providing customers with a robust tool for making well-informed decisions, we hope to curb instances of overcharging and deceptive practices.

- By mitigating scams and malpractices, our project not only contributes to the evolution of predictive modeling in India's automotive industry but also promotes transparency and fairness, fostering positive change for consumers.

OBJECTIVE

The objective of the ‘Car Quintessence’ project is to develop a user-friendly web application that leverages a trained Random Forest Regressor model.
The application allows users to input car parameters, such as Km driven, model, year, fuel type, etc and receive accurate predictions of car prices. Utilizing a dataset collected from reputable sources, the goal is to provide individuals with a simple and effective tool for estimating car values based on machine learning insights, enhancing the user experience in the process.

INTRODUCTION

CHALLENGES IN AUTOMOTIVE VALUATION : In the realm of automotive valuation, a persistent challenge has lingered – the lack of a precise and transparent mechanism for predicting car prices. Traditional methods fell short in capturing the nuanced factors influencing prices, paving the way for our project to address this critical gap.

HISTORICAL INEFFICIENCIES : The inadequacy of the older methods prompted our foray into the development of a sophisticated predictive model, poised to overcome the inefficiencies of the past.

THE VOID IN TRANSPARENCY : A prevailing issue in the automotive sector has been the lack of transparency in pricing, leading to instances of overcharging and unfair practices. Our project emerges as a solution to this problem, aiming to introduce a level of transparency that empowers consumers and instills trust in the valuation process.

CLOSING THE PRECISION GAP : Previous models struggled to achieve the desired precision in predicting car prices, often resulting in suboptimal decision-making for consumers. Recognizing this gap, our project sets out to redefine precision in automotive valuation, offering a solution that goes beyond conventional limitations.

ENVIRONMENTAL IMPACT : Beyond predicting car prices, our project acknowledges the global imperative for sustainability. By integrating considerations for eco-friendly vehicles, we promote informed choices that contribute to a reduced carbon footprint and a more sustainable automotive future. Through encouraging the adoption of environmentally conscious vehicles, we aim to play a role in mitigating the environmental impact of the automotive industry.

VISION FOR FUTURE & OUR SOLUTION : our project envisions a future where predictive modeling becomes synonymous with accuracy, transparency, and fairness in the automotive valuation landscape. We embark on this journey to revolutionize the field, providing a solution that not only addresses historical inefficiencies but sets new standards for predictive modeling in the automotive industry.

LITERATURE SURVEY

Title: Predicting the price of used cars using machine learning techniques.

Author: Pudaruth

Year: 2017

Description: According to a World Economic Forum's report, AI-enabled automation will generate 133 million new jobs globally by 2022. And in India itself, the demand for AI talent pool is expected to skyrocket with the government's steps towards digitization, and multiple organizations accelerating their digital transformation initiatives. Are you ready to ride the wave? The BITS Pilani 11-month online PG Programme in AI & ML is designed to help working professionals like you to develop an understanding of AI & ML and its various building blocks.

Title: PIN Optimal Distribution of Auction Vehicles System: Applying Price Forecasting, Elasticity Estimation, and Genetic Algorithms to Used-Vehicle Distribution. Marketing Science

Year: 2009

Author: Du, J., Xie, L., & Schroeder

Description: In addition to retailing new vehicles, automotive manufacturers in the United States sell millions of vehicles through leasing and to fleet customers every year. The majority of these vehicles are returned to the automotive manufacturers at the end of the contracted term and must be "remarketed." In 2007, about 10 million used vehicles were sold at more than 400 auctions in the United States. Large consigners face decisions every day about when, where, and at what price to offer these vehicles, which has significant financial implications for their profitability. To address the challenges of the distribution process, (), a division of J.D. Power and Associates, developed the PIN Optimal Distribution of Auction Vehicles System (ODAV).

EXISTING SYSTEM

In the existing system, shop owners are implementing a car price prediction system using the k-means algorithm. K-means is a popular clustering algorithm that helps identify groups or clusters within a dataset based on similarities in features. By applying the k-means algorithm to a dataset of car attributes such as brand, model, year, mileage, and condition, the system can segment cars into distinct groups. This clustering enables the system to analyze the characteristics of each group and predict the price range for a given car based on its attributes.

DISADVANTAGES OF EXISTING SYSTEM

- Less accuracy
- Difficult to predict
- With global clusters, it doesn’t work well

PROPOSED SYSTEM

In this system, we implement an effective car price prediction system using the Random Forest Algorithm, offering a seamless user experience with both CSV file uploads and manual entries. Beyond its standalone capabilities, our system features web integration, allowing users to access the prediction service through a user-friendly web interface. This integration enhances accessibility, enabling users to input car parameters effortlessly and receive accurate predictions in real-time. Whether through direct data upload or interactive manual entry, the system's web integration ensures a versatile and dynamic platform for efficient car valuation.

ADVANTAGES OF EXISTING SYSTEM

- Easy to predict the car prices
- High accuracy rate (over 90 %)
- Enhanced Precision and Accuracy
- User-Friendly Web Integration
- Adaptability to Diverse Data Sources
- Real-Time Predictions
- Transparency and Trust

MINIMUM SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:-

Processor : Intel Core i3-8100 (Quad Core CPU or Higher)

GPU : NVIDIA GeForce GTX 1000 series CUDA-enabled GPU for faster model training (Optional but recommended for large datasets)

RAM : 8 GB DDR4 2666Mt/s or higher

Storage : 256 GB SATA HDD at 7200 RPM (preferably PCIe Gen3 SSD for faster data access)

SOFTWARE REQUIREMENTS:-

Operating System : Windows 7 or MacOS X 10.13 or Ubuntu 18.04 or higher

IDE : Microsoft Visual Studio / Anaconda / Jupyter / Spyder, etc (IDE of choice)

Python Environment : Python3 (version 3.9.18)

Libraries for Data Processing : Pandas (version 2.1.3) , NumPy (version 1.26.2)

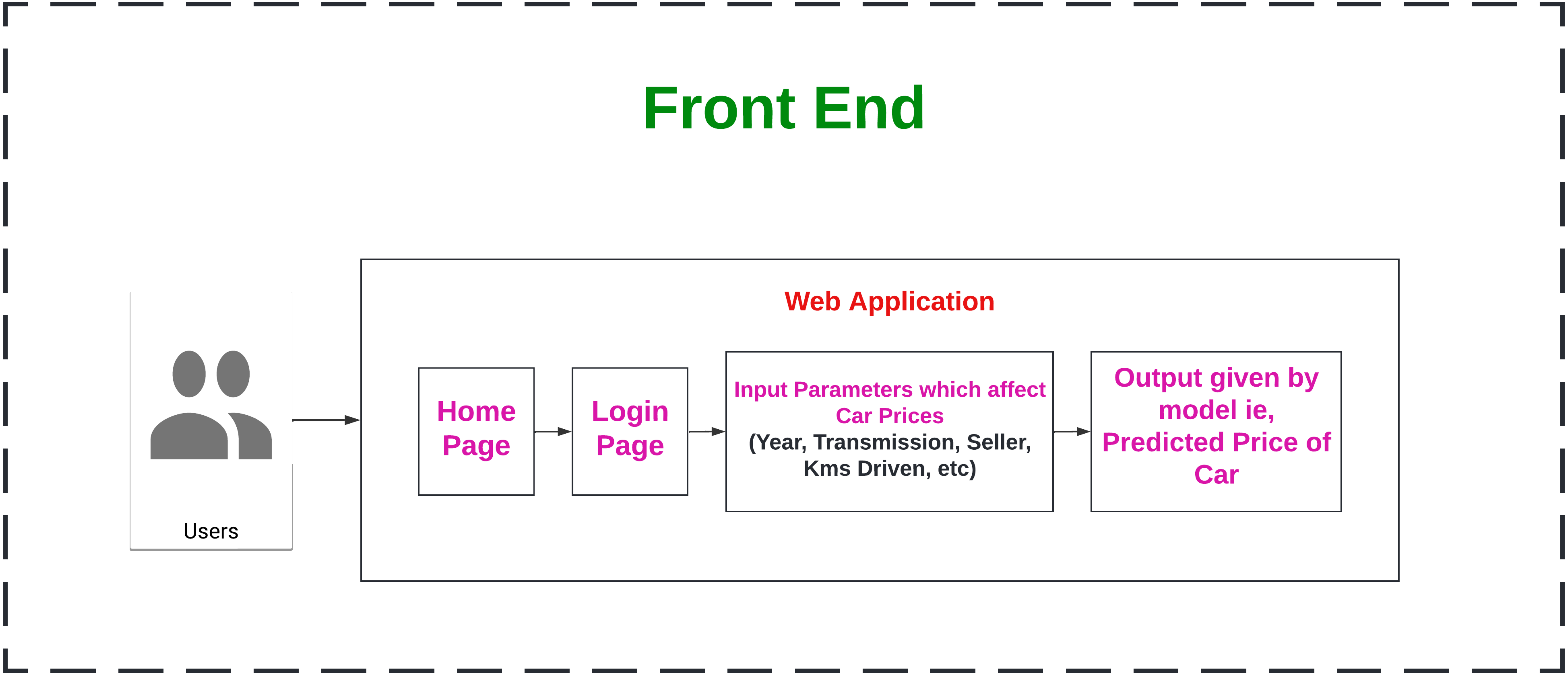
Machine Learning Libraries : Scikit-learn (version 0.23.2 or 0.24.2

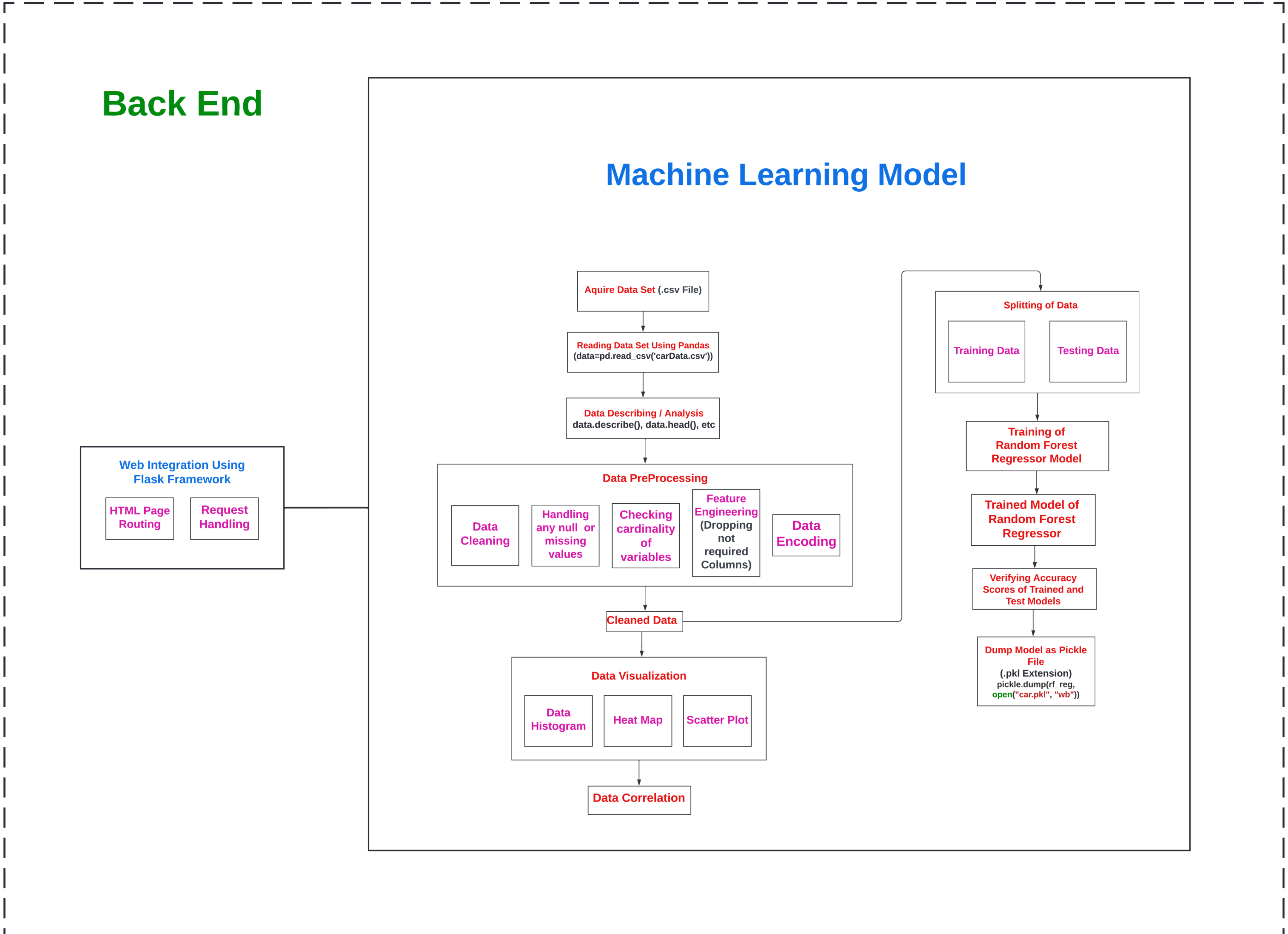
Database : Local Storage (No Separate Database)

Framework : Flask (version 2.2.2)

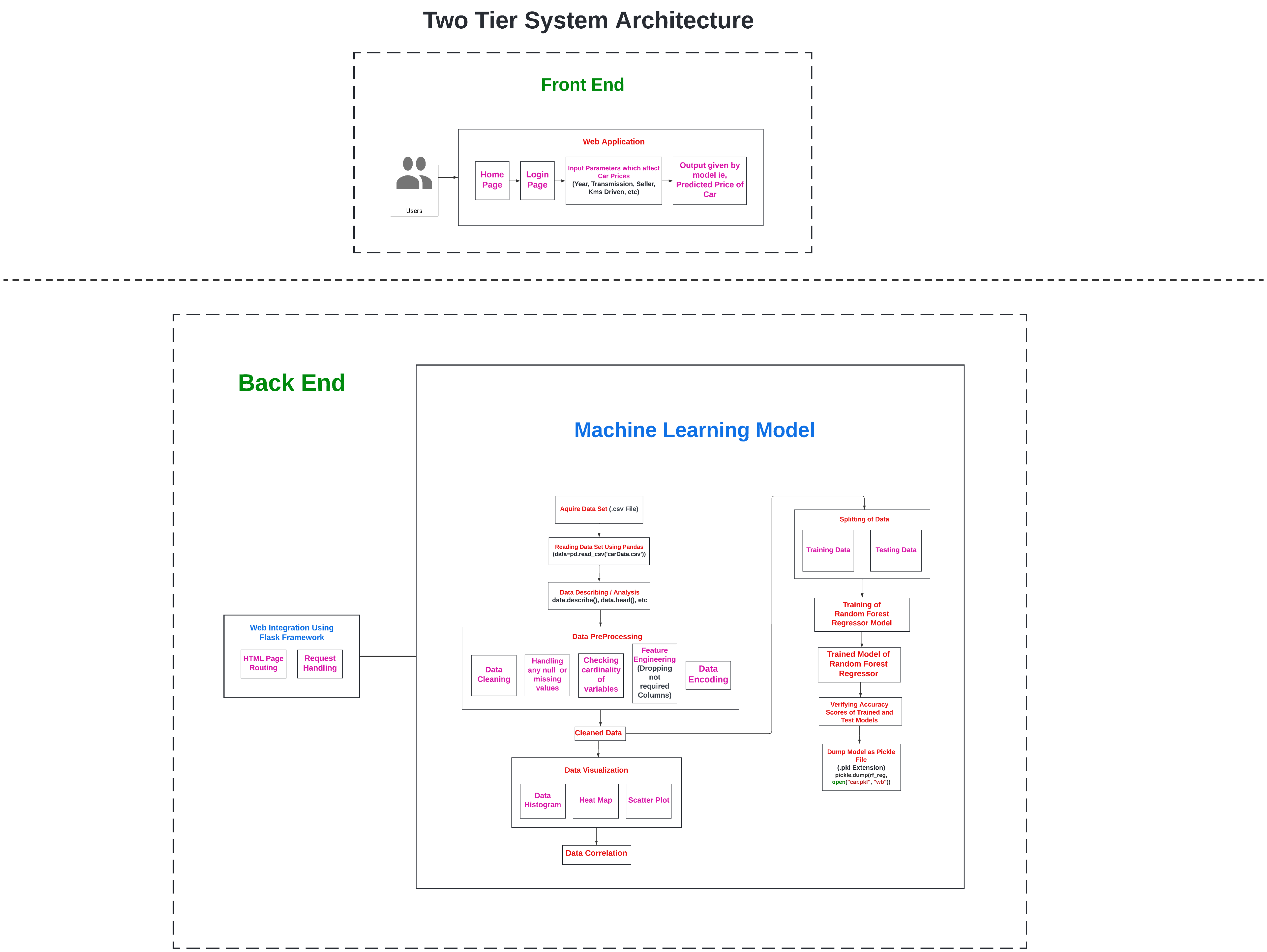
Frontend (Web Application) : HTML5, CSS3, JavaScript ver ES6 (Preferably Google Chrome Browser)

System Architecture



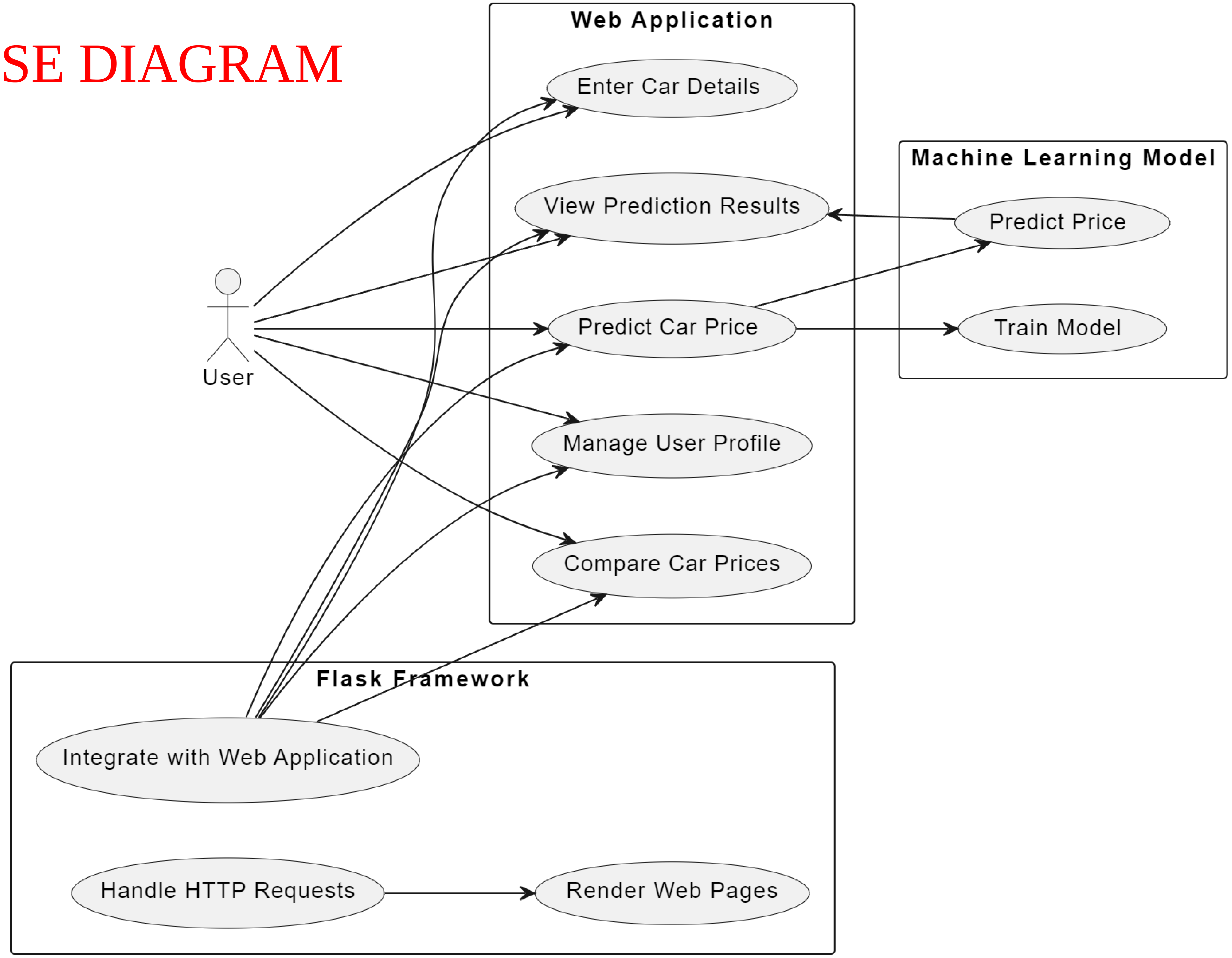


Back End
Machine Learning Model
Web Integration Using Flask Framework
HTML Page Routing
Request Handling
Aquire Data Set (.csv File)
Reading Data Set Using Pandas (data=pd.read_csv('carData.csv'))
Data Describing / Analysis data.describe(), data.head(), etc
Data PreProcessing
Data Cleaning
Handling any null or missing values
Checking cardinality of variables
Feature Engineering (Dropping not required Columns)
Data Encoding
Cleaned Data
Data Visualization
Data Histogram
Heat Map
Scatter Plot
Data Correlation
Splitting of Data
Training Data
Testing Data
Training of Random Forest Regressor Model
Trained Model of Random Forest Regressor
Verifying Accuracy Scores of Trained and Test Models
Dump Model as Pickle File (.pkl Extension) pickle.dump(rf_reg, open("car.pkl", "wb"))

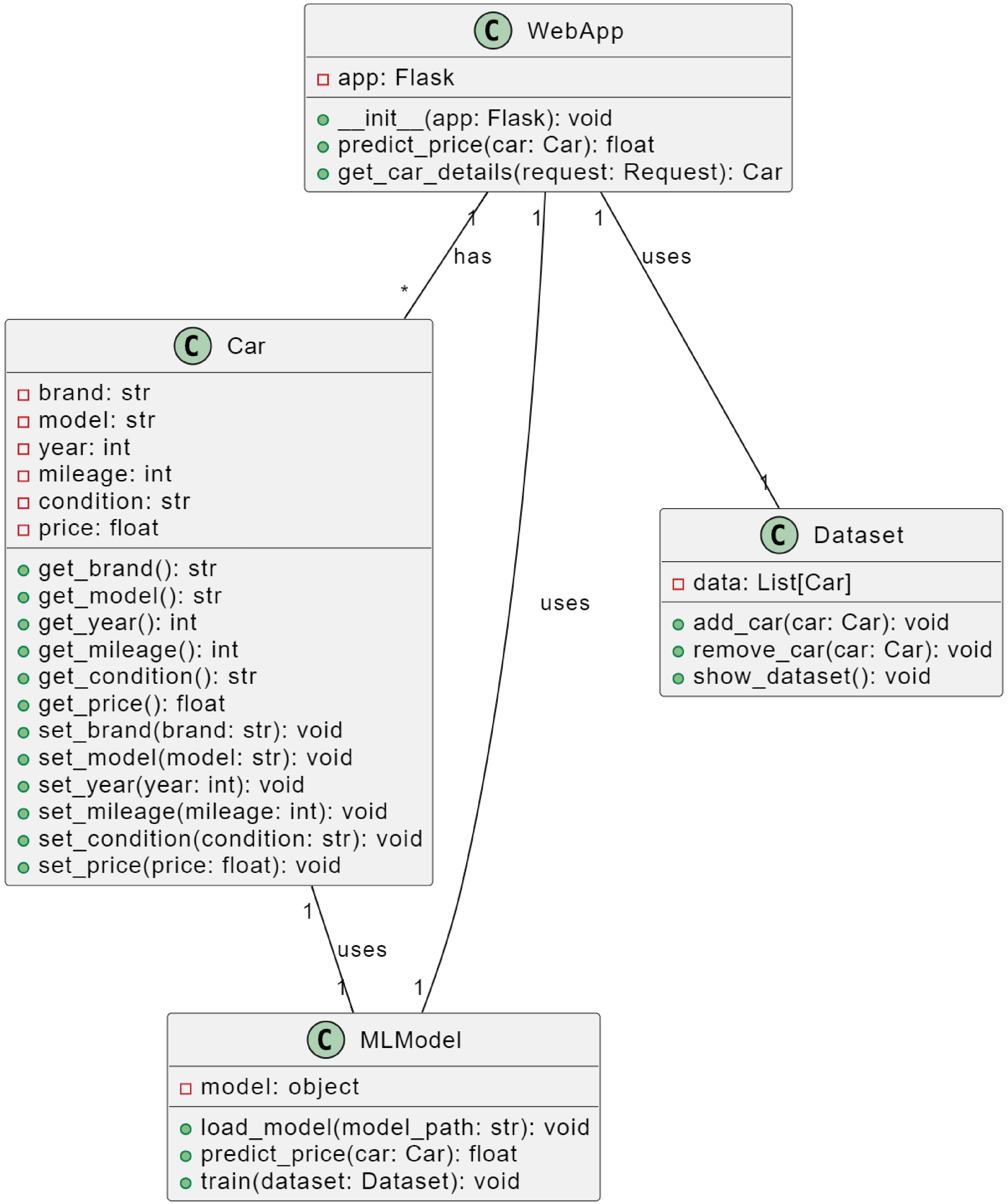

Two Tier System Architecture
Front End
Web Application
Users
Home Page
Login Page
Input Parameters which affect Car Prices
(Year, Transmission, Seller, Kms Driven, etc)
Output given by model ie, Predicted Price of Car
Back End
Web Integration Using Flask Framework
HTML Page Routing
Request Handling
Machine Learning Model
Aquire Data Set (.csv File)
Reading Data Set Using Pandas
(data=pd.read_csv('carData.csv'))
Data Describing / Analysis
data.describe(), data.head(), etc
Data PreProcessing
Data Cleaning
Handling any null or missing values
Checking cardinality of variables
Feature Engineering (Dropping not required Columns)
Data Encoding
Cleaned Data
Data Visualization
Data Histogram
Heat Map
Scatter Plot
Data Correlation
Splitting of Data
Training Data
Testing Data
Training of Random Forest Regressor Model
Trained Model of Random Forest Regressor
Verifying Accuracy Scores of Trained and Test Models
Dump Model as Pickle File
(.pkl Extension)
pickle.dump(rf_reg, open("car.pkl", "wb"))

System Design (UML Diagrams)

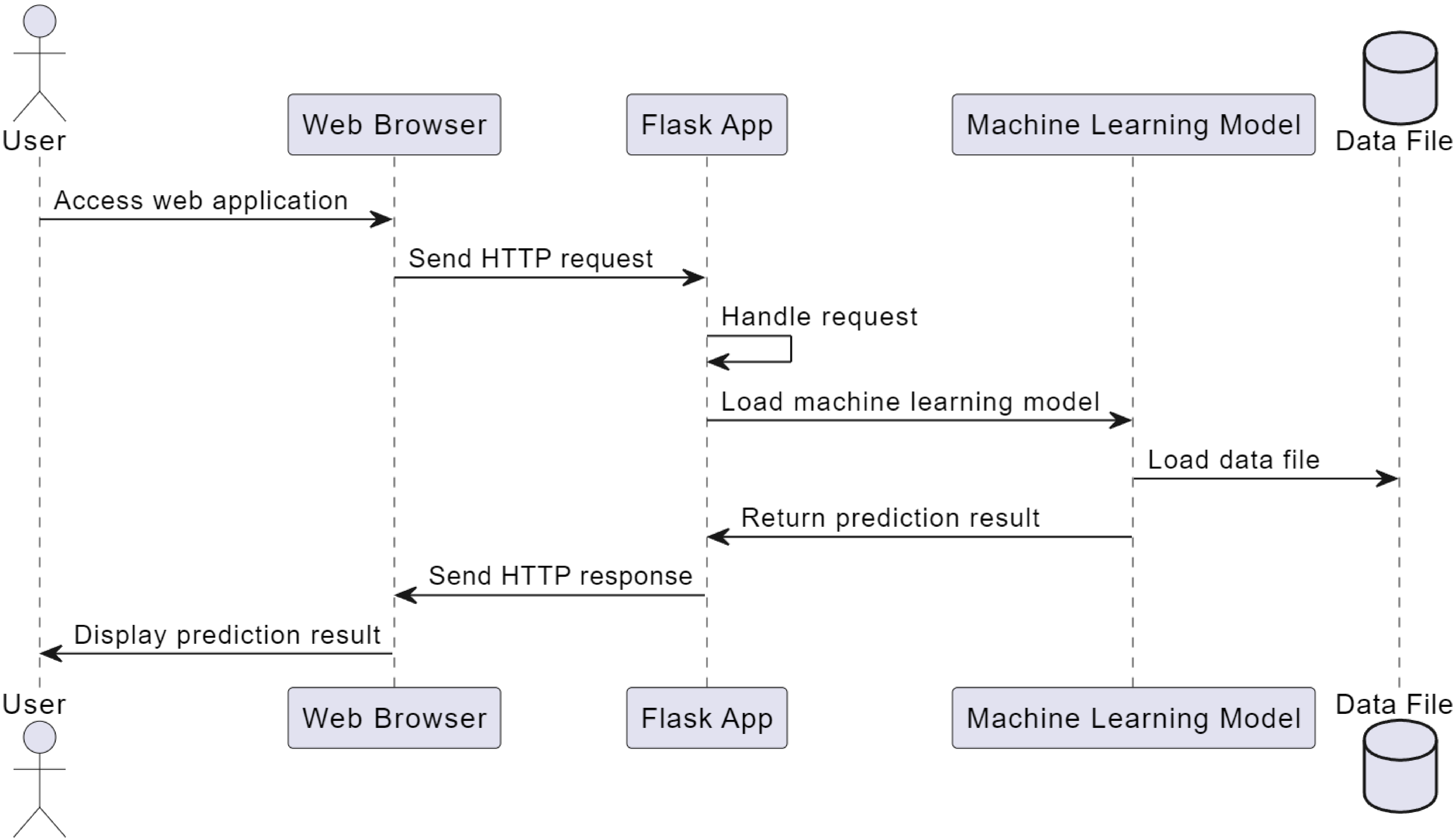
1. USE CASE DIAGRAM



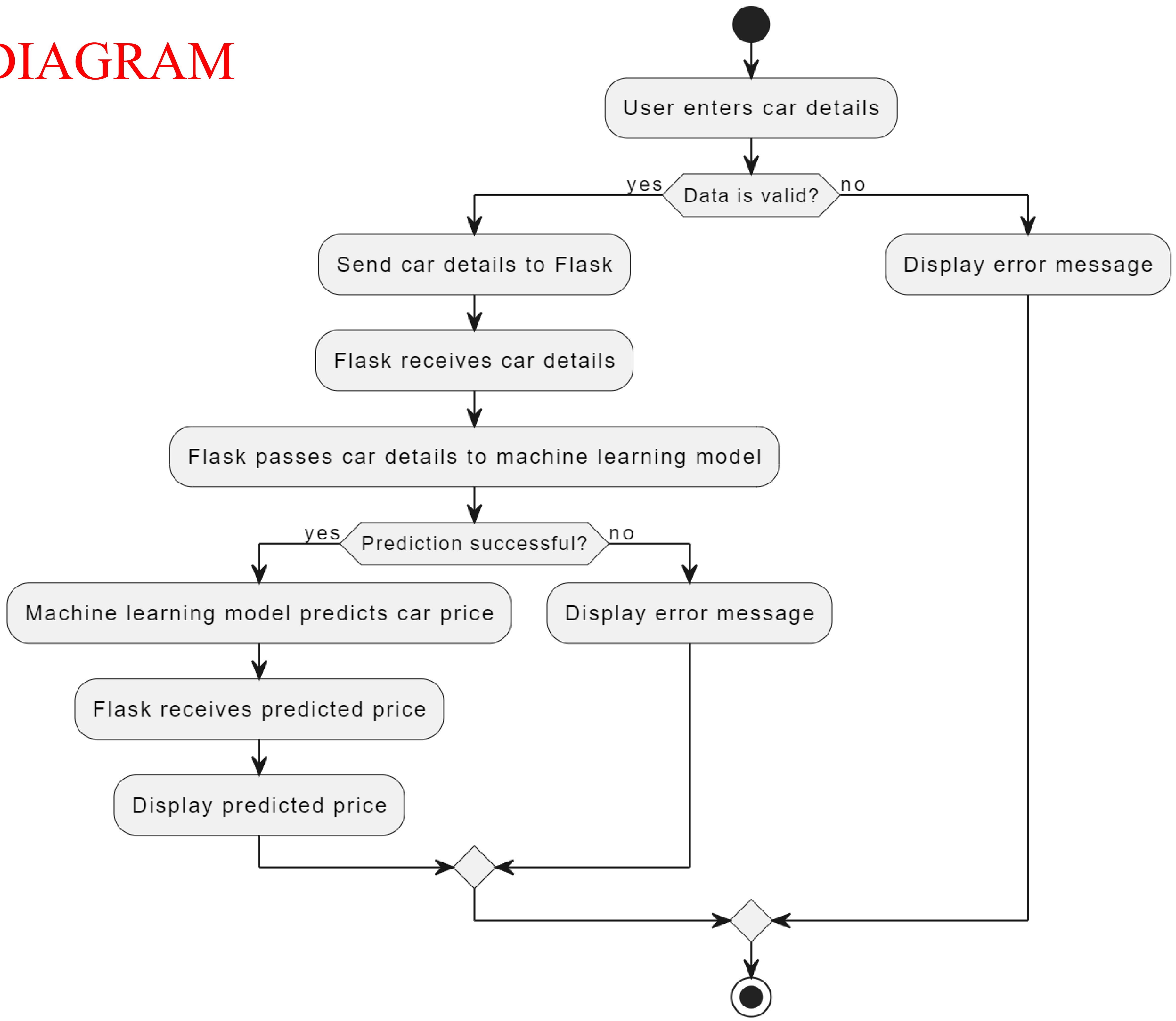
2. Class Diagram



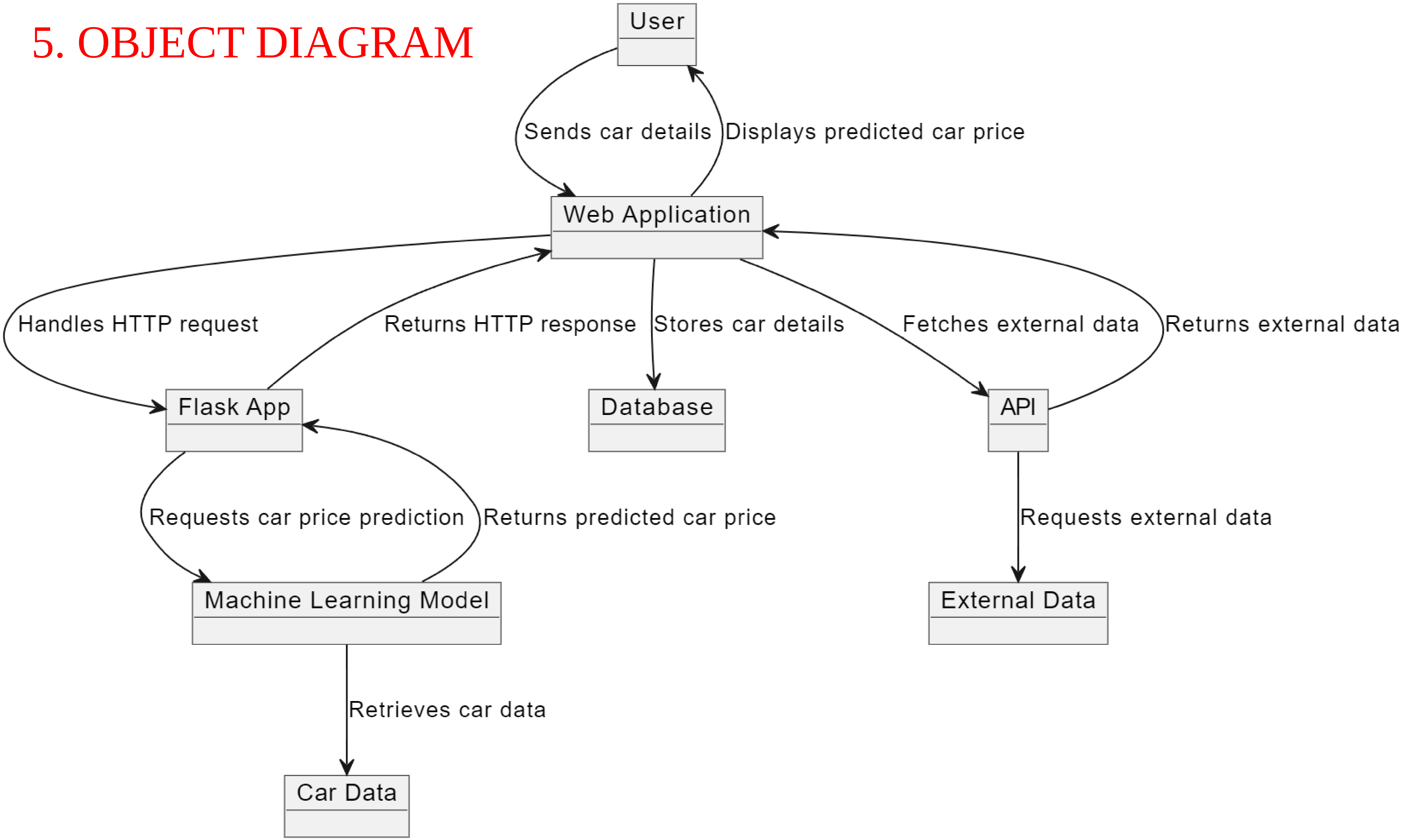
3. SEQUENCE DIAGRAM



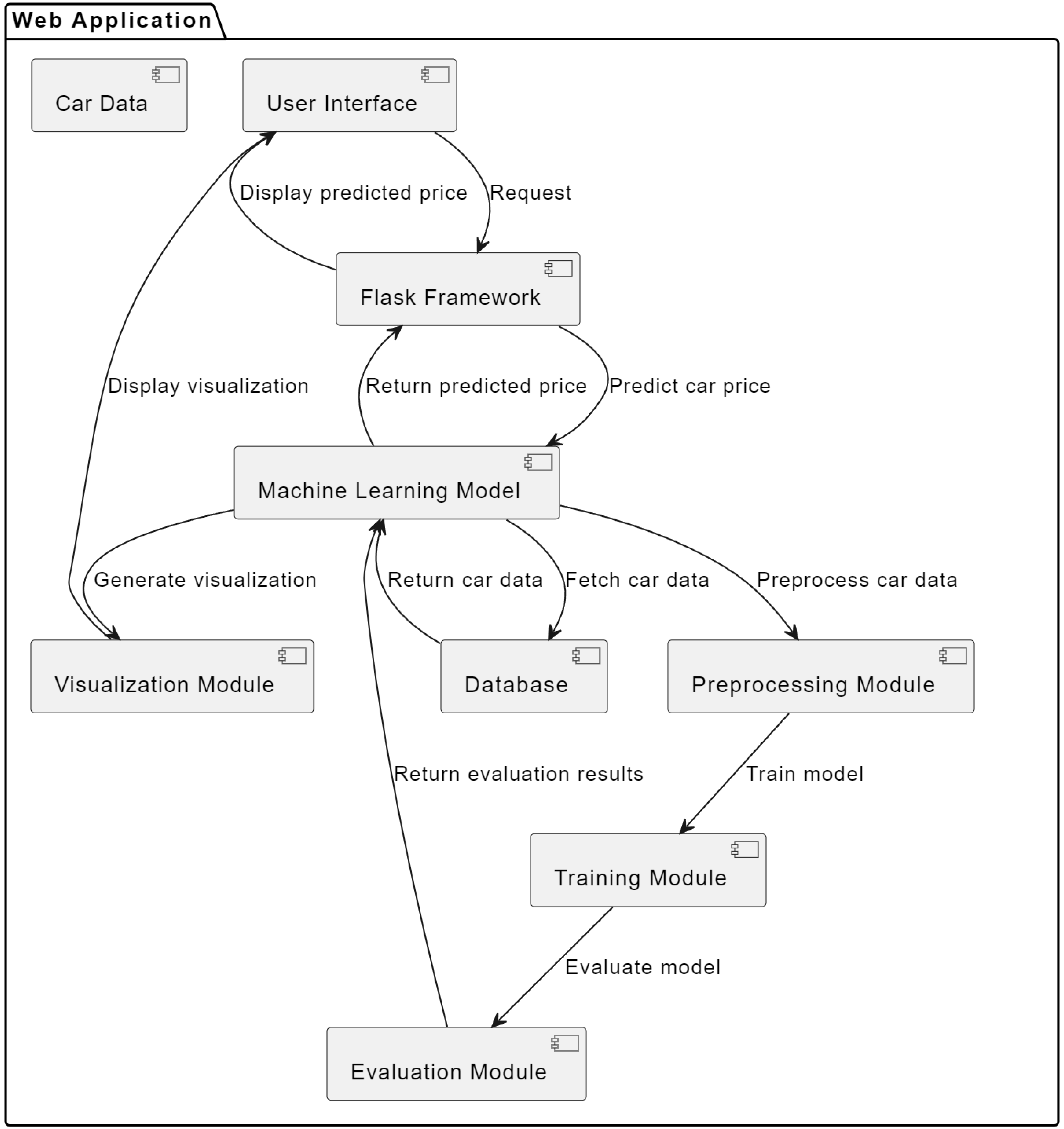
4. ACTIVITY DIAGRAM



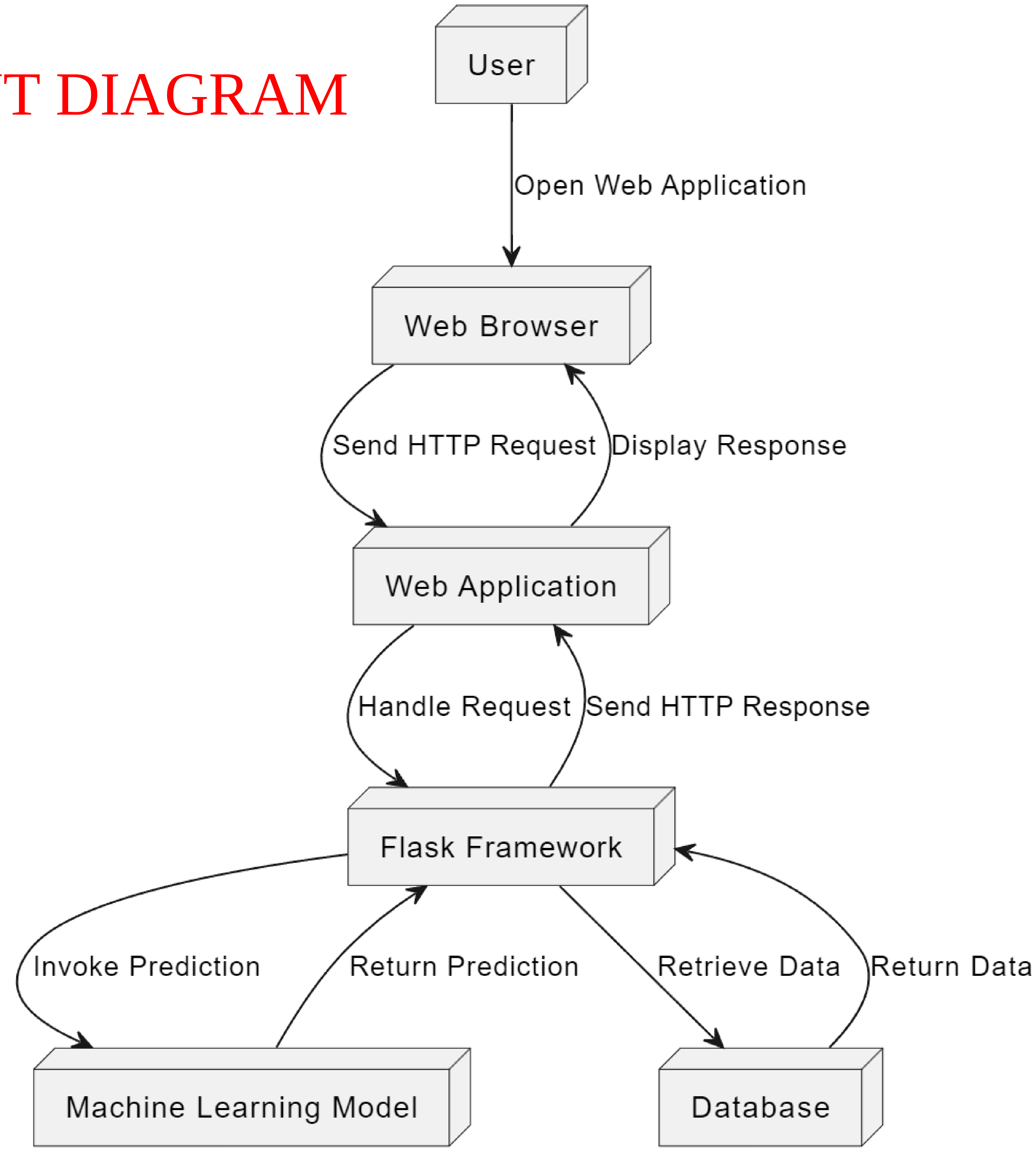
5. OBJECT DIAGRAM



6. COMPONENT DIAGRAM

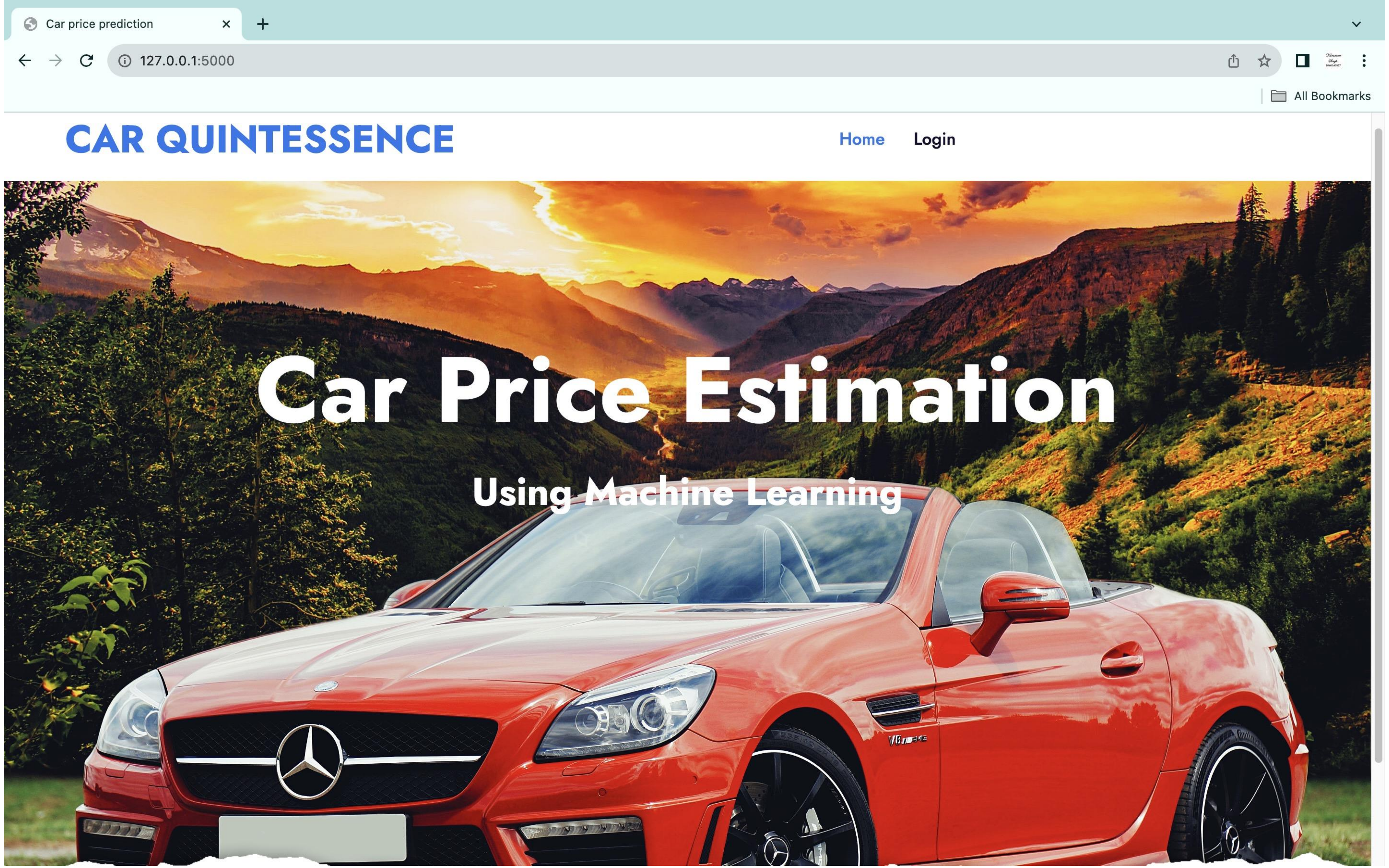


7. DEPLOYMENT DIAGRAM

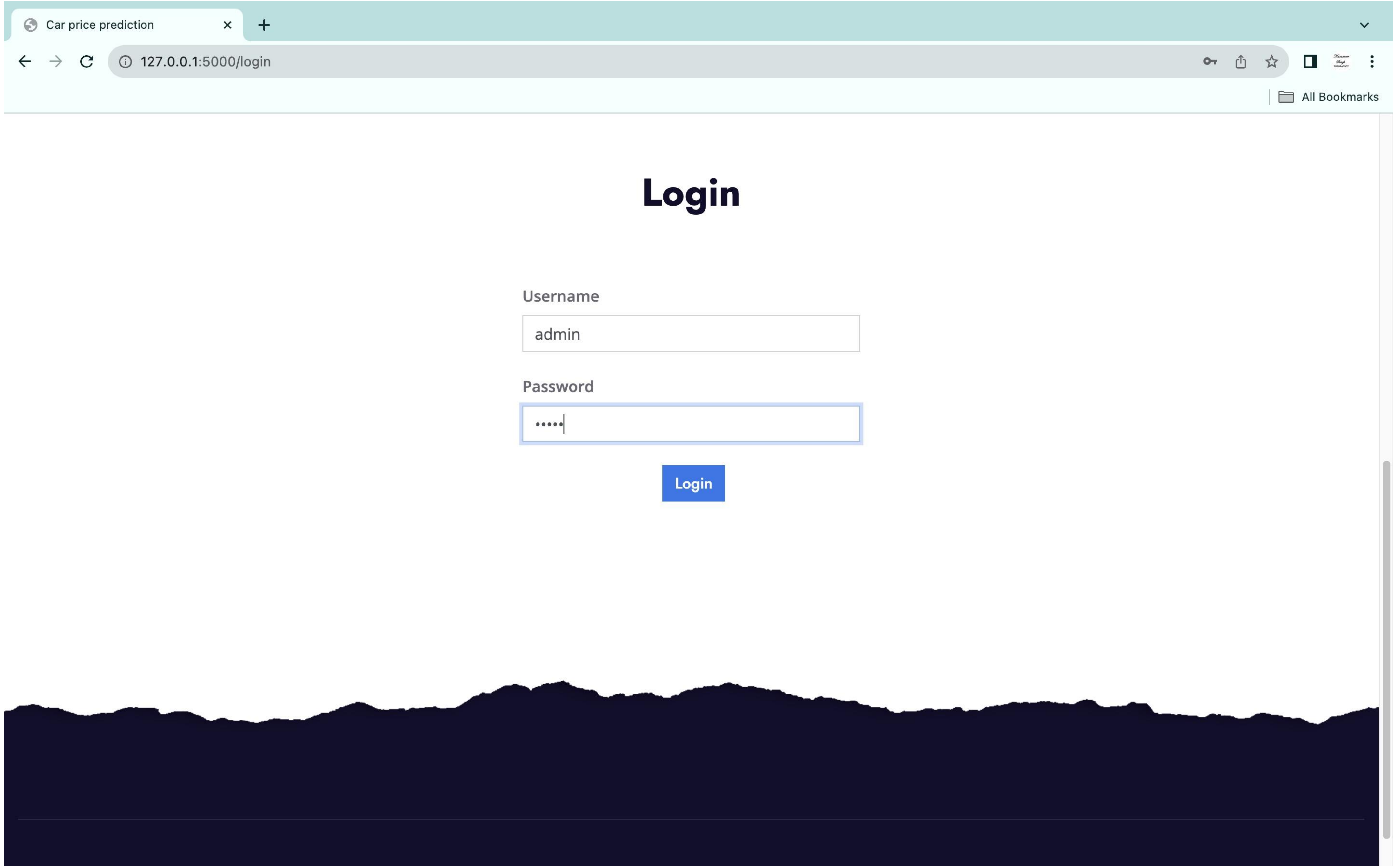


SCREEN SHOTS

1. HOME PAGE

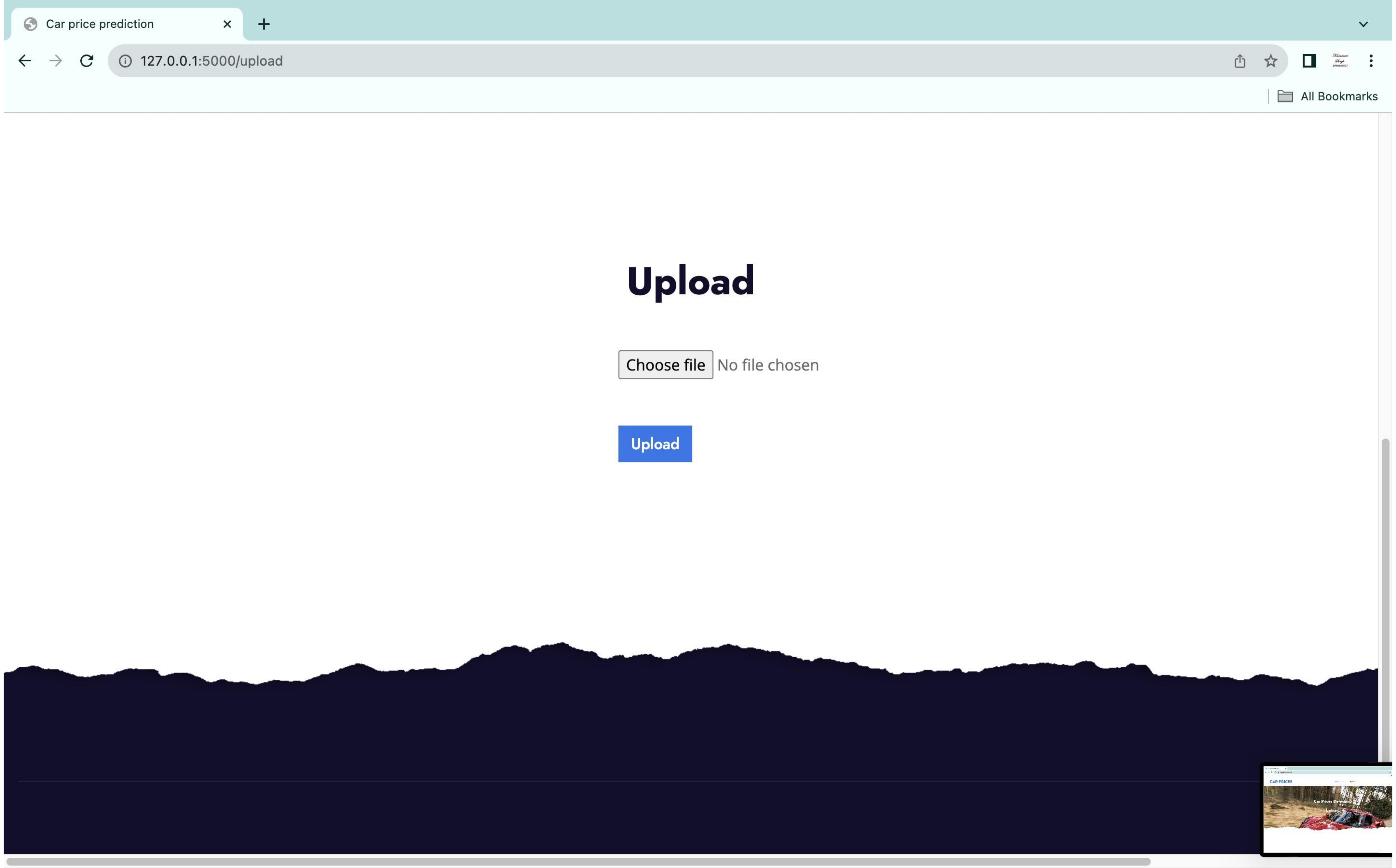


2. LOGIN PAGE



Car price prediction
127.0.0.1:5000/login
All Bookmarks
Login
Username
admin
Password
Login

3. UPLOAD .CSV FILE



4. PREVIEW DATA FRAME

Car price prediction

localhost:5000/preview

All Bookmarks

Preview

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
Id									
1	ritz	2014	3.35	5.590	27000	Petrol	Dealer	Manual	0
2	sx4	2013	4.75	9.540	43000	Diesel	Dealer	Manual	0
3	ciaz	2017	7.25	9.850	6900	Petrol	Dealer	Manual	0
4	wagon r	2011	2.85	4.150	5200	Petrol	Dealer	Manual	0
5	swift	2014	4.60	6.870	42450	Diesel	Dealer	Manual	0
6	vitara brezza	2018	9.25	9.830	2071	Diesel	Dealer	Manual	0
7	ciaz	2015	6.75	8.120	18796	Petrol	Dealer	Manual	0
8	s cross	2015	6.50	8.610	33429	Diesel	Dealer	Manual	0
9	ciaz	2016	8.75	8.890	20273	Diesel	Dealer	Manual	0
10	ciaz	2015	7.45	8.920	42367	Diesel	Dealer	Manual	0
11	alto 800	2017	2.85	3.600	2135	Petrol	Dealer	Manual	0

5. TRAIN THE MACHINE LEARNING MODEL

Car price prediction

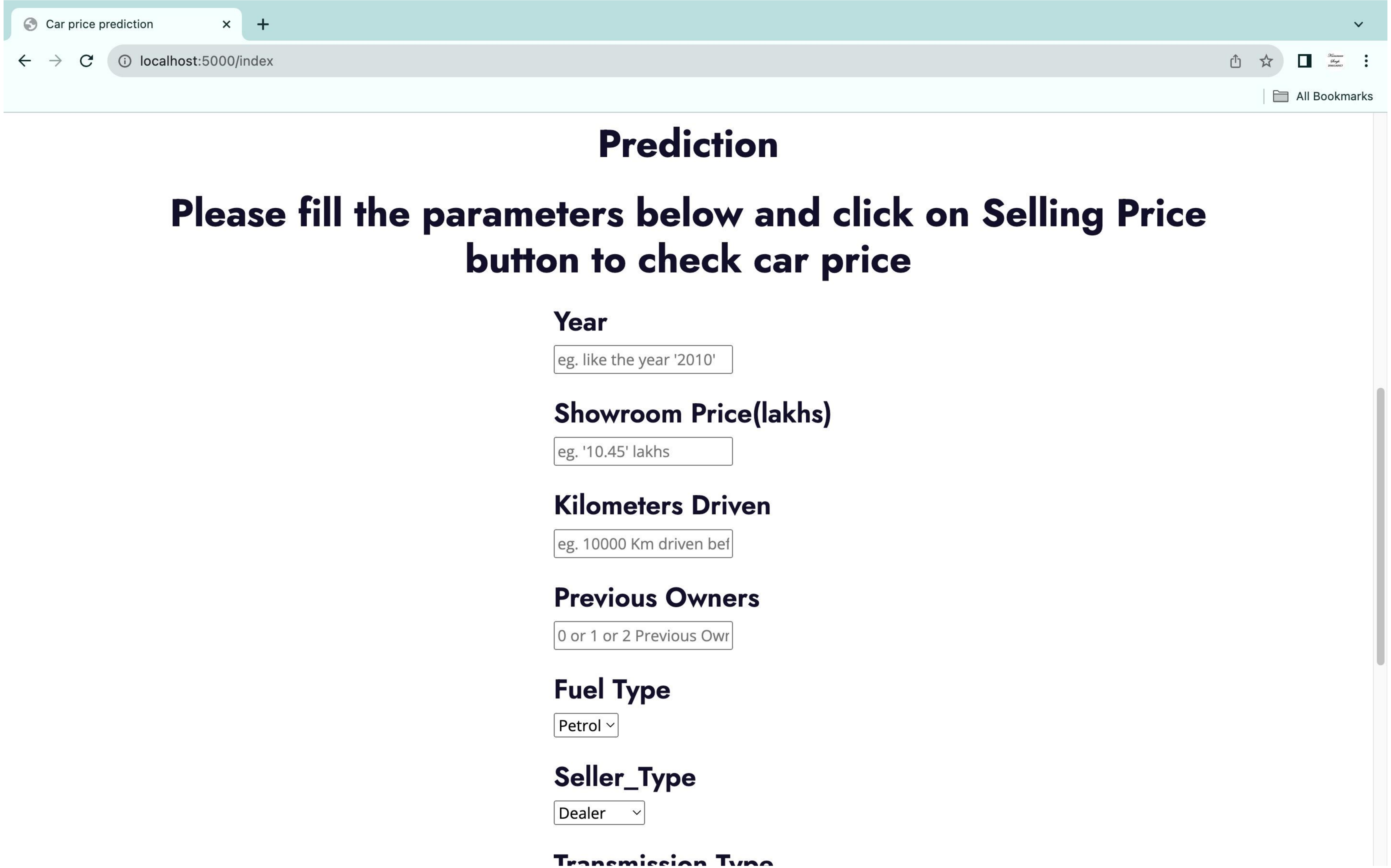
localhost:5000/preview

All Bookmarks

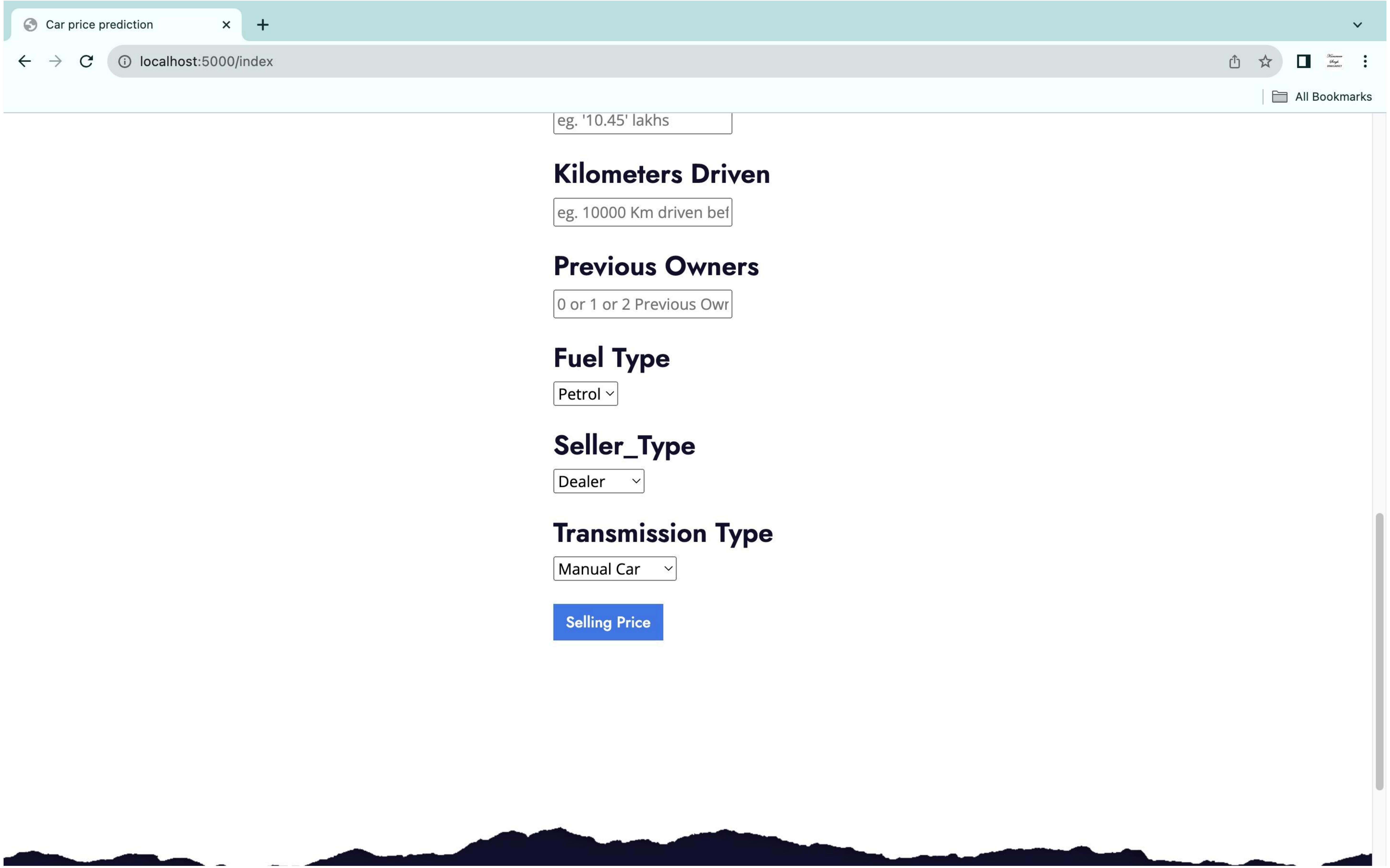
292	brio	2015	5.40	6.100	31427	Petrol	Dealer	Manual	0
293	jazz	2016	6.40	8.400	12000	Petrol	Dealer	Manual	0
294	city	2010	3.25	9.900	38000	Petrol	Dealer	Manual	0
295	amaze	2014	3.75	6.800	33019	Petrol	Dealer	Manual	0
296	city	2015	8.55	13.090	60076	Diesel	Dealer	Manual	0
297	city	2016	9.50	11.600	33988	Diesel	Dealer	Manual	0
298	brio	2015	4.00	5.900	60000	Petrol	Dealer	Manual	0
299	city	2009	3.35	11.000	87934	Petrol	Dealer	Manual	0
300	city	2017	11.50	12.500	9000	Diesel	Dealer	Manual	0
301	brio	2016	5.30	5.900	5464	Petrol	Dealer	Manual	0

Click to Train | Test

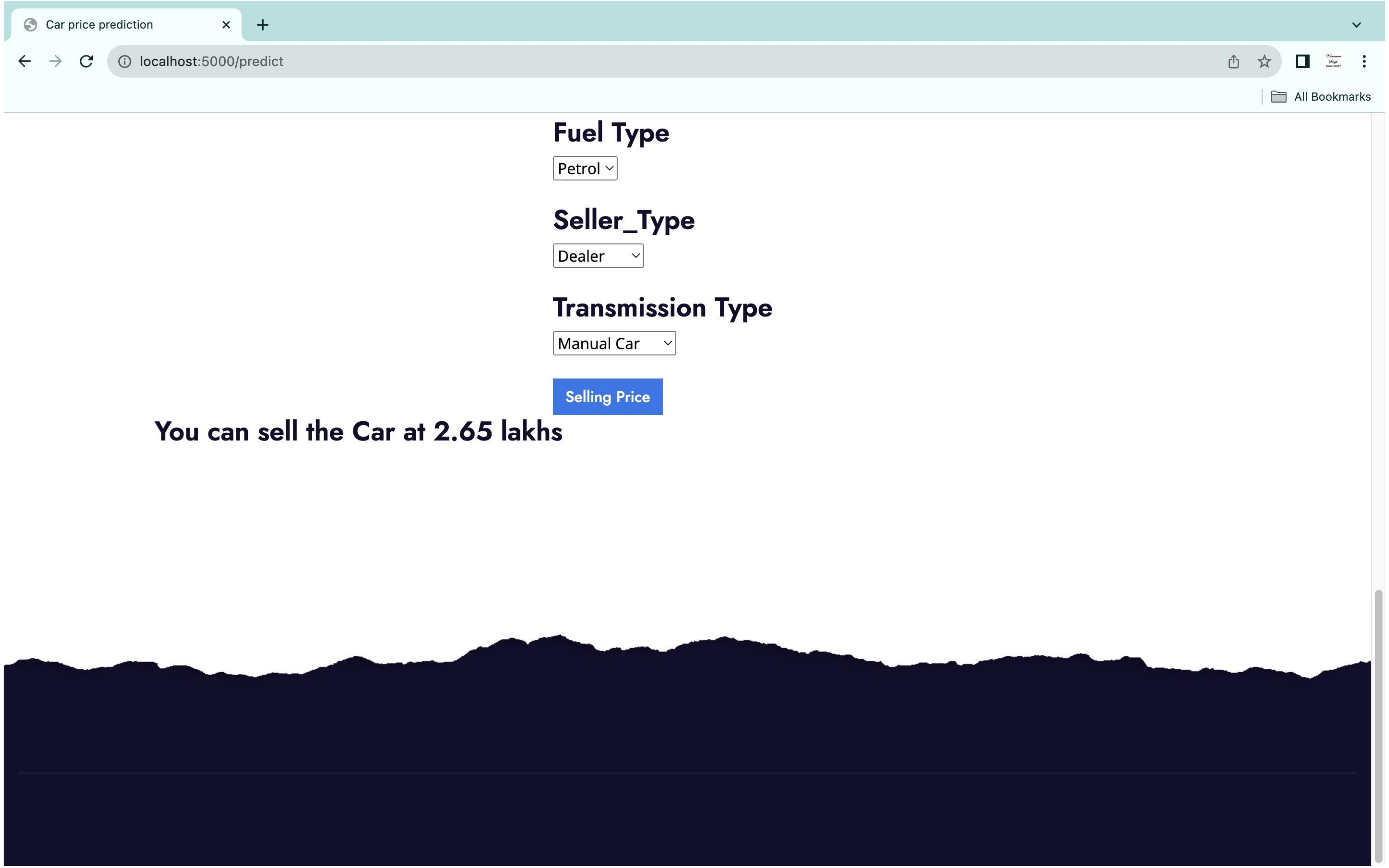
6. ENTERING CAR PARAMETERS



7. ENTERING CAR FEATURES



8. PREDICTED CAR PRICE (OUTPUT)



9. PERFORMANCE OF THE MODEL

Car price prediction

localhost:5000/performance

PERFORMANCE ANALYSIS

The r2_score: 0.92

Mean Absolute Error: 0.81

RMSE: 1.54

Standard Deviation: 9.22 %

CONCLUSION

Car price prediction poses a formidable challenge due to the multitude of attributes crucial for accurate estimations. A pivotal phase in this process involves data collection and preprocessing. In this research, data normalization, standardization, and cleaning were executed through a dedicated script, mitigating unnecessary noise for machine learning algorithms. Although data cleaning significantly enhances prediction performance, it proved insufficient for the complexities inherent in the dataset. Employing other machine learning algorithms resulted in an accuracy below 50%. Consequently, this project proposes the Random Forest Algorithm Regression approach, achieving a notable accuracy of ~92%. This marked improvement underscores the effectiveness of this approach. However, it is acknowledged that this system demands more computational resources than other machine learning algorithms.

REFERENCES

[1] A. Das Mou, P. K. Saha, S. A. Nisher and A. Saha, "A Comprehensive Study of Machine Learning algorithms for Predicting Car Purchase Based on Customers Demands," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 2021,IEEE, pp. 180-184, doi: 10.1109/ICICT4SD50815.2021.9396868.

[2] J. C. Pope and J. Silva-Risso, "The psychological effect of weather on car purchases* meghan r. busse devin g. pope," The Quarterly Journal of Economics, vol. 1, no. 44, p. 44, 2014.

[3] M. Jayakameswaraiah and S. Ramakrishna, "Development of data mining system to analyze cars using tknn clustering algorithm," In- ternational Journal of Advanced Research in Computer Engineering Technology, vol. 3, no. 7, 2014.

[4] F.Harahap,A.Y.N.Harahap,E.Ekadiansyah,R.N.Sari,R.Adawiyah, and C. B. Harahap, "Implementation of na ̈ive bayes classification method for predicting purchase," in 2018 6th International Conference on Cyber and IT Service Management (CITSM). IEEE, 2018, pp. 1–5.

[5] K. S. Durgesh and B. Lekha, "Data classification using support vector machine," Journal of theoretical and applied information technology, vol. 12, no. 1, pp. 1–7, 2010.

[6] R. Ragupathy and L. Phaneendra Maguluri, "Comparative analysis of machine learning algorithms on social media test," International Journal of Engineering and Technology(UAE), vol. 7, pp. 284–290, 03 2018.

[7] N.Pal,P.Arora,P.Kohli,D.Sundararaman,andS.S.Palakurthy,“How much is my car worth? a methodology for predicting used cars prices using random forest,” in Future of Information and Communication Conference. Springer, 2018, pp. 413–422.

[8] S. Pudaruth, “Predicting the price of used cars using machine learning techniques,” Int. J. Inf. Comput. Technol, vol. 4, no. 7, pp. 753–764, 2014.

[9] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, “Supervised machine learning algorithms: classification and comparison,” International Journal of Computer Trends and Technology (IJCTT), vol. 48, no. 3, pp. 128–138, 2017.

[10] M. R. Busse, D. G. Pope, J. C. Pope, and J. Silva-Risso, “The psychological effect of weather on car purchases,” The Quarterly Journal of Economics, vol. 130, no. 1, pp. 371–414, 2015.

[11] S. Veni and A. Srinivasan, “Defect classification using na ̈ive bayes classification,” Interbational Journal of Applied Engineering Research, vol. 12, no. 22, pp. 12 693–12 700, 2017.

[12] E. Gegic, B. Isakovic, D. Keco, Z. Masetic, and J. Kevric, “Car price prediction using machine learning techniques,” 2019.

[13] M. Jabbar, “Prediction of heart disease using k-nearest neighbor and particle swarm optimization,” Biomed. Res, vol. 28, no. 9, pp. 4154– 4158, 2017.

FUTURE ENHANCEMENTS

Although, this system has achieved astonishing performance in car price prediction problem our aim for the future research is the following:-

DYNAMIC PRICE PREDICTION MODELS:

Implement dynamic pricing models that consider real-time market trends, economic factors, and emerging technologies, providing users with even more accurate and up-to-date car price predictions.

USER CUSTOMIZATION AND PREFERENCES:

Introduce user profiles where individuals can set preferences, allowing the system to tailor predictions based on specific criteria like brand loyalty, environmental consciousness, or desired features.

INCORPORATION OF EXTERNAL DATASETS:

Enhance the dataset by integrating external sources such as automotive industry reports, consumer sentiment data, and environmental impact assessments to further refine predictions and provide a holistic view of the market.

INTERACTIVE DATA VISUALIZATIONS:

Develop interactive visualizations that allow users to explore and interpret the factors influencing car prices, promoting a deeper understanding of the prediction model and fostering user engagement.

GEO-SPECIFIC PRICE INSIGHTS:

Provide geographically specific insights into regional pricing variations, considering factors like local demand, supply chain dynamics, and economic conditions to offer users location-specific predictions.

INTEGRATION OF ADVANCED MACHINE LEARNING ALGORITHMS:

Explore and integrate advanced machine learning algorithms beyond Random Forest, such as neural networks or gradient boosting, to evaluate their impact on prediction accuracy and enhance the system's capabilities.

MOBILE APPLICATION INTEGRATION:

Develop a mobile application to extend the accessibility of the car price prediction system,enabling users to input parameters and receive predictions conveniently on their smartphones.

HISTORICAL PRICE TRENDS ANALYSIS:

Implement a feature that analyzes historical car price trends, allowing users to track how prices have evolved over time and providing valuable insights for predicting future trends.

THANK YOU