USED CAR PRICE PREDICTION SYSTEM

-REPORT BY VIRESH ACHARYA

Date: 26-9-23

Abstract:

The AI product we propose is an AI-driven car price prediction model designed to cater to the automotive industry. The primary aim is to assist potential car buyers, car dealerships, and insurers in predicting the most accurate and fair market value of used cars. This product utilizes machine learning algorithms and data analysis to generate predictions based on various car features, historical sales data, and market trends. By providing reliable car price estimates, the AI model aims to enhance the car buying and selling experience and optimize pricing strategies for businesses in the automotive market.

1. Problem Statement:

The automotive industry faces challenges in accurately determining the fair market value of used cars due to various factors like age, mileage, condition, and market fluctuations. Car buyers often find it difficult to assess whether the quoted price is reasonable, while dealerships struggle to optimize pricing to attract potential customers. An AI-powered car price prediction model can address this problem by providing data-driven and precise car valuations.

The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But, due to the increased prices of new cars and the financial incapability of the 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 which effectively determines the worthiness of the car using a variety of features. Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and its value in the present-day scenario. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at. To overcome this problem, we have developed a model which will be highly effective. Machine learning Algorithms are used because they provide us with continuous value as an output and not a categorized value. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.

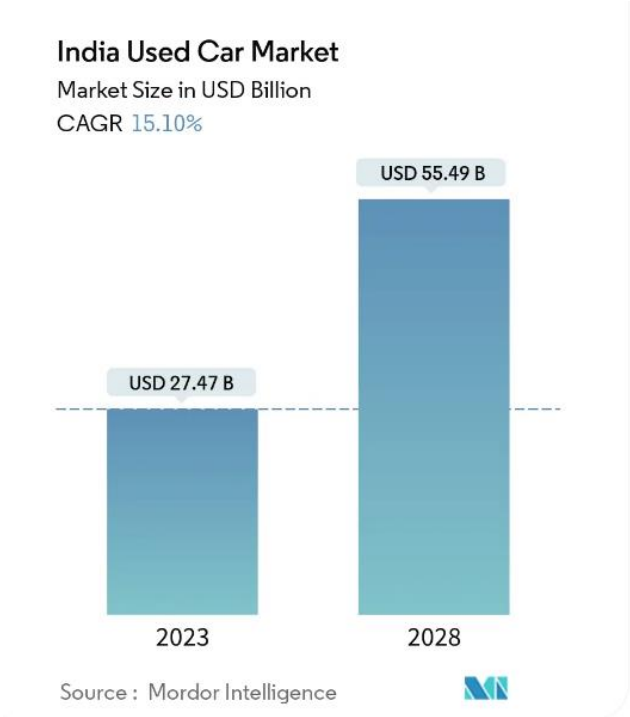
2. Market/Customer/Business Need Assessment:

The target market for this AI product includes car dealerships, used car sellers, insurance companies, and individual car buyers. Car dealerships can utilize this tool to set competitive prices and enhance their sales strategy. Insurance companies can incorporate this model to assess the accurate value of insured vehicles, preventing overvaluation or undervaluation. Individual car buyers can leverage the AI model to make informed purchasing decisions.

2.1 India Used Car Market Analysis

The India Used Car Market size is expected to grow from USD 27.47 billion in 2023 to USD 55.49 billion by 2028, at a CAGR of 15.10% during the forecast period (2023-2028).

The COVID-19 pandemic had a minimal impact on the industry. With the increased number of people preferring individual mobility and more finance options available in the used car market, the market is set to grow considerably. Due to the pandemic's reduced cash inflow, buyers have been forced to look for alternatives to new cars, and the used car industry has high growth potential in this regard. As the pandemic has hampered new vehicle sales and production, the used car market is gaining traction among buyers.



Study Period	2018-2027
Base Year For Estimation	2022
Market Size (2023)	USD 27.47 Billion
Market Size (2028)	USD 55.49 Billion
CAGR (2023 - 2028)	15.10 %
Market Concentration	Medium

Major Players

*Disclaimer: Major Players sorted in no particular order

3.Target Specifications and Characterization:

The ideal customer profile for our AI product includes car dealerships and used car sellers who want to optimize pricing strategies, insurance companies seeking accurate valuation for insured vehicles, and individual car buyers looking for fair market value estimations.

The AI model for car price prediction targets the automotive industry and various stakeholders within it. The primary customers and users of this AI product include:

Car Dealerships: Car dealerships require an accurate and data-driven car price prediction system to set competitive prices for their inventory. They aim to attract potential buyers and optimize their pricing strategies to increase sales and revenue.

Used Car Sellers: Individual sellers or small-scale used car dealers can benefit from the AI model to determine the fair market value of their vehicles. This helps them avoid overpricing or underpricing and improves their chances of making successful sales.

Insurance Companies: Insurance companies often need to assess the accurate value of insured vehicles, especially in the case of insurance claims or policy renewals. The AI model provides an objective estimation, ensuring fair compensation and preventing overvaluation.

Car Buyers: Individual car buyers seeking to purchase a used car can use the AI model to evaluate whether the quoted price is reasonable and aligns with the car's features, condition, and market trends. This empowers buyers to make informed decisions and enhances their overall buying experience.

4. External Search

The sources I have used as reference for analyzing the need of such a system for Car Prediction Model have mentioned below :

- Price Prediction of Used Cars Using Machine Learning
- India Used Car Market Analysis
- Used Cars Research
- Kaggle car price prediction

5.Benchmarking:

Benchmarking alternate products of car price predictions involves comparing the features, performance, and effectiveness of existing solutions available in the market.

Traditional Appraisal Methods: Traditional methods involve manual appraisal by experienced car appraisers or dealers who consider factors like the car's age, mileage, condition, market demand, and historical sales data. Benchmarking against these methods can help assess the accuracy and efficiency of AI-driven predictions.

Online Valuation Tools: Various online platforms offer car valuation services where users can input car details, and the system provides an estimated price. Benchmarking against these tools will help determine how AI-powered predictions compare in terms of accuracy and ease of use.

Data-Driven Pricing Platforms: Some companies offer data-driven pricing solutions that leverage historical sales data and market trends to set optimal car prices. Benchmarking against these platforms will reveal the effectiveness of AI models in predicting prices based on historical data.

Human Expertise: Car dealers and experts with years of experience in the automotive industry often rely on their intuition and knowledge to price used cars. Benchmarking AI predictions against human expertise will provide insights into AI model accuracy and potential areas of improvement.

Auction Websites: Online car auction platforms provide a glimpse of how much buyers are willing to pay for a specific car model. Benchmarking AI predictions against actual auction results will assess how well the AI model captures market demand.

Private Sales Data: Some sellers prefer private sales through classified ads or personal networks. Comparing AI predictions with actual private sale prices will gauge the model's accuracy in real-world scenarios.

Dealership Pricing: Analyzing car prices set by dealerships for similar models can help benchmark AI predictions against market standards and dealership practices.

Price Comparison Websites: Websites that aggregate car prices from various sources provide valuable data for benchmarking AI predictions against market trends and competitor pricing.

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6019 entries, 0 to 6018
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Unnamed: 0         6019 non-null   int64  
 1   Name               6019 non-null   object 
 2   Location           6019 non-null   object 
 3   Year               6019 non-null   int64  
 4   Kilometers_Driven  6019 non-null   int64  
 5   Fuel_Type          6019 non-null   object 
 6   Transmission       6019 non-null   object 
 7   Owner_Type         6019 non-null   object 
 8   Mileage            6017 non-null   object 
 9   Engine             5983 non-null   object 
 10  Power              5983 non-null   object 
 11  Seats              5977 non-null   float64
 12  New_Price          824 non-null    object 
 13  Price              6019 non-null   float64
dtypes: float64(2), int64(3), object(9)
memory usage: 658.5+ KB
```

6.Applicable Regulations:

While car price prediction AI models primarily deal with data analysis and prediction, there are certain regulations that need to be considered when developing and deploying such systems. The applicable regulations can vary depending on the country and region. Here are some key regulations that may be relevant to consider for a car price prediction AI model:

6.1 Data Protection and Privacy Regulations: Any AI model that deals with customer data, including car owners' information, must comply with data protection and privacy laws. For example, the European Union's General Data Protection Regulation (GDPR) imposes strict rules on the collection, storage, and processing of personal data.

6.2 Fair Trade and Consumer Protection Laws: Car price prediction models should adhere to fair trade practices and comply with consumer protection laws to ensure transparency and avoid unfair pricing practices.

6.3 Anti-Discrimination Laws: AI models must not discriminate against individuals based on protected characteristics such as race, gender, age, or ethnicity. The model should be designed to ensure fairness and avoid biased predictions.

6.4 Intellectual Property Rights: Car price prediction models should not infringe on existing patents or intellectual property rights of other companies or individuals.

6.5 Export and Import Regulations: If the AI model is deployed in multiple countries, it is essential to comply with export and import regulations for technology and software.

7.Applicable Constraints

Data Availability: The availability and quality of historical car sales data and relevant features (make, model, year, mileage, condition, etc.) play a significant role in training an accurate AI model. Limited or low-quality data may impact the model's performance.

Computational Resources: Developing and training AI models can require substantial computational power, especially for complex machine learning algorithms. Constraints related to hardware and processing capabilities need to be considered.

Model Complexity: The complexity of the AI model should be balanced with the available data and computational resources. Overly complex models may lead to overfitting and make it challenging to interpret the results.

Expertise and Skillsets: Developing a car price prediction AI model requires expertise in machine learning, data science, and software development. Constraints related to available expertise and skillsets should be taken into account.

Model Maintenance and Updates: AI models require continuous monitoring, maintenance, and updates to remain accurate and relevant over time. Constraints related to ongoing support and maintenance need to be considered.

Deployment Environment: The AI model's deployment environment, such as on-premises or cloud-based, can affect factors like scalability, reliability, and data privacy.

Cost and Budget: The development and deployment of AI models can incur costs related to data acquisition, computational resources, software, and personnel. Constraints related to the available budget should be carefully managed.

Ethical Considerations: The AI model should be designed with ethical considerations in mind, avoiding biased predictions or discriminatory outcomes.

Scalability: The AI model should be scalable to handle an increasing number of car listings and predictions as the user base grows.

User Acceptance and Trust: The AI model's predictions should be accurate and trustworthy to gain user acceptance and build trust in the system.

8. Business Opportunity:

The car price prediction AI model presents a significant business opportunity in the automotive industry. By leveraging machine learning algorithms and data analysis techniques, this AI-driven solution can provide valuable insights and benefits to various stakeholders. Here are some key aspects of the business opportunity:

Improved Pricing Strategies: Car dealerships can use the AI model to set competitive and data-driven prices for their inventory. By considering factors such as historical sales data, market trends, and car features, dealers can optimize pricing strategies to attract potential buyers and increase sales.

Enhanced Customer Experience: Car buyers seeking used cars can benefit from the AI model's accurate price predictions. Transparent and fair pricing fosters trust and enhances the overall customer experience, increasing the likelihood of successful sales.

Cost Savings: The AI model can help individual sellers or small-scale used car dealers avoid overpricing or under-pricing their vehicles. By selling cars at fair market values, sellers can optimize their profits and avoid potential financial losses.

Efficient Insurance Claims: Insurance companies can use the AI model to assess the accurate value of insured vehicles. In the case of insurance claims or policy renewals, this ensures fair compensation and prevents overvaluation, ultimately leading to more efficient claims processing.

Data Monetization: If the AI model collects anonymized data from user interactions, car manufacturers or other automotive industry players can leverage this data for market analysis, trend identification, and data-driven decision-making.

9.Concept Generation

Identify the Problem: Clearly define the problem we want to solve with the car price prediction AI model. For example, the problem could be providing accurate and data-driven car price estimates to help sellers and buyers make informed decisions.

Define the Target Audience: Determine the target audience for the AI model. It could be car dealerships, individual sellers, insurance companies, or car buyers.

Understand User Needs: Conduct market research and gather feedback from potential users to understand their pain points, requirements, and expectations from the AI model.

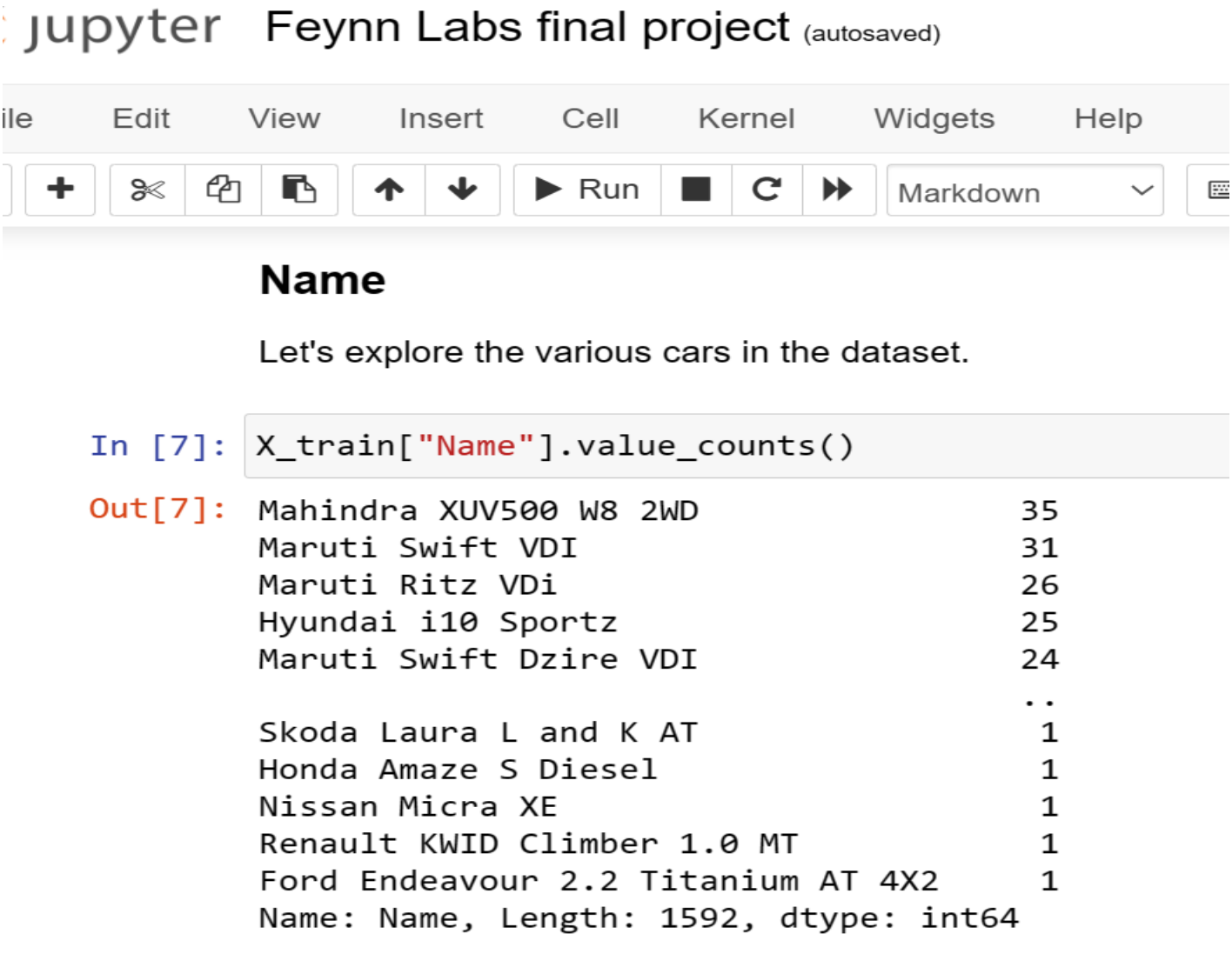
Explore Existing Solutions: Analyze existing car price prediction tools and solutions to identify their strengths, weaknesses, and gaps that your AI model can address.

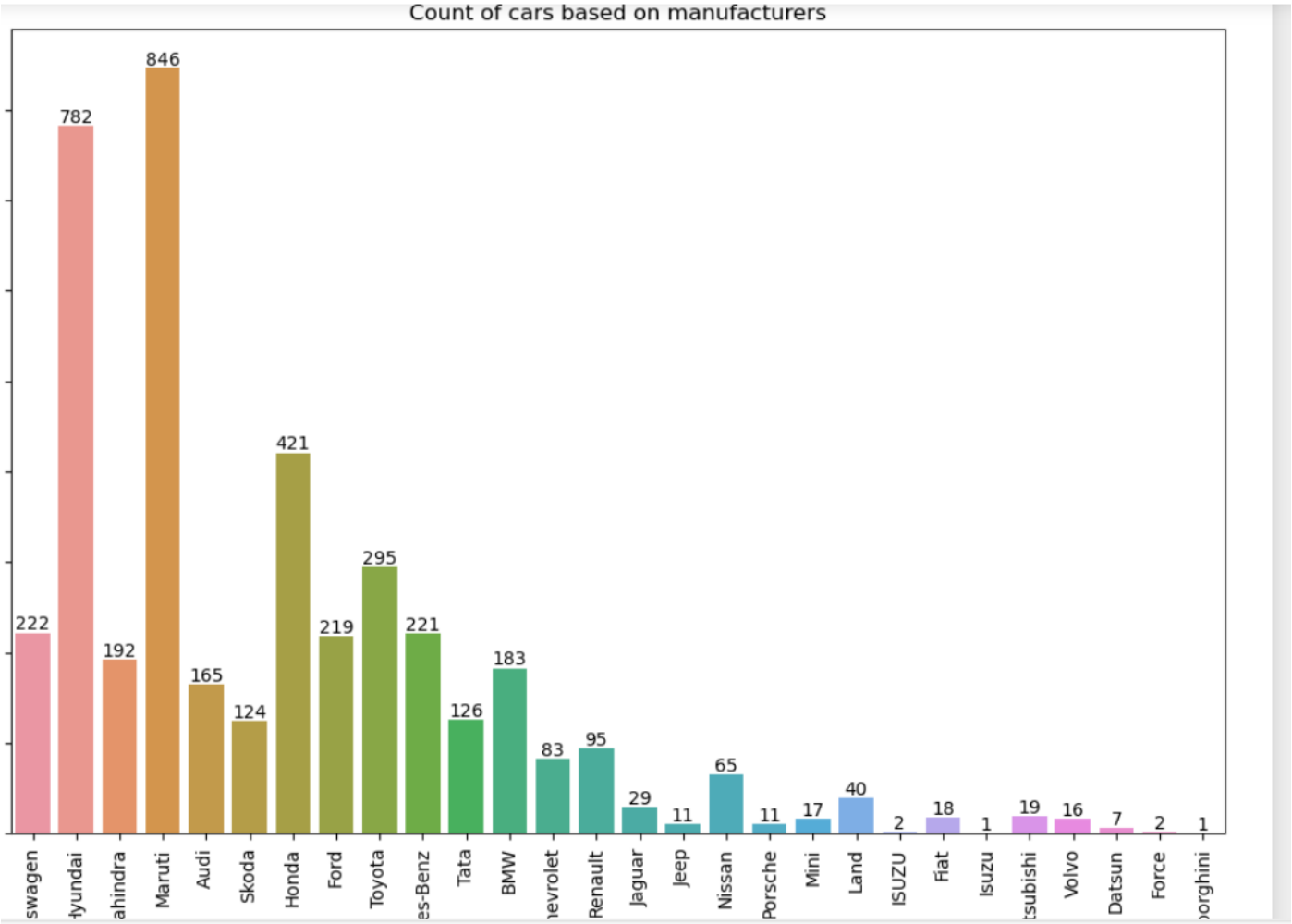
Determine Key Features: Based on user needs and the identified problem, list down the key features that the car price prediction AI model should have. This could include factors like car make, model, year, mileage, condition, market trends, and image analysis.

Choose Machine Learning Algorithms: Select appropriate machine learning algorithms for regression and predictive modeling. Consider algorithms such as linear regression, decision tree regression, or more advanced methods like neural networks.

Data Sources and Collection: Identify relevant data sources, including historical car sales data, market trends, and car specifications. Plan for data collection and preprocessing to ensure data quality and consistency.

CODE IMPLEMENTATION (SMALL SCALE):





Data Processing

Now that we have worked with the training data, let's create dummy columns for categorical columns before we begin training.

```
X_train = pd.get_dummies(X_train,
                         columns = ["Manufacturer", "Fuel_Type", "Transmission",
                         drop_first = True)
```

```
X_test = pd.get_dummies(X_test,
                         columns = ["Manufacturer", "Fuel_Type", "Transmission",
                         drop_first = True)
```

Simple Linear Regression

Simple Linear Regression is a method to help us understand the relationship between two variables:

- The predictor/independent variable (X)
- The response/dependent variable (that we want to predict)(Y)

The result of Linear Regression is a **linear function** that predicts the response (dependent) variable as a function of the predictor (independent) variable.

$$Y : Response\ Variable$$
$$X : Predictor\ Variables$$

Linear Function

$$Yhat = a + bX$$

- a refers to the **intercept** of the regression line, in other words: the value of Y when X is 0
- b refers to the **slope** of the regression line, in other words: the value with which Y changes when X increases by 1 unit

Intercept and coefficient

```
linearRegression.intercept_
```

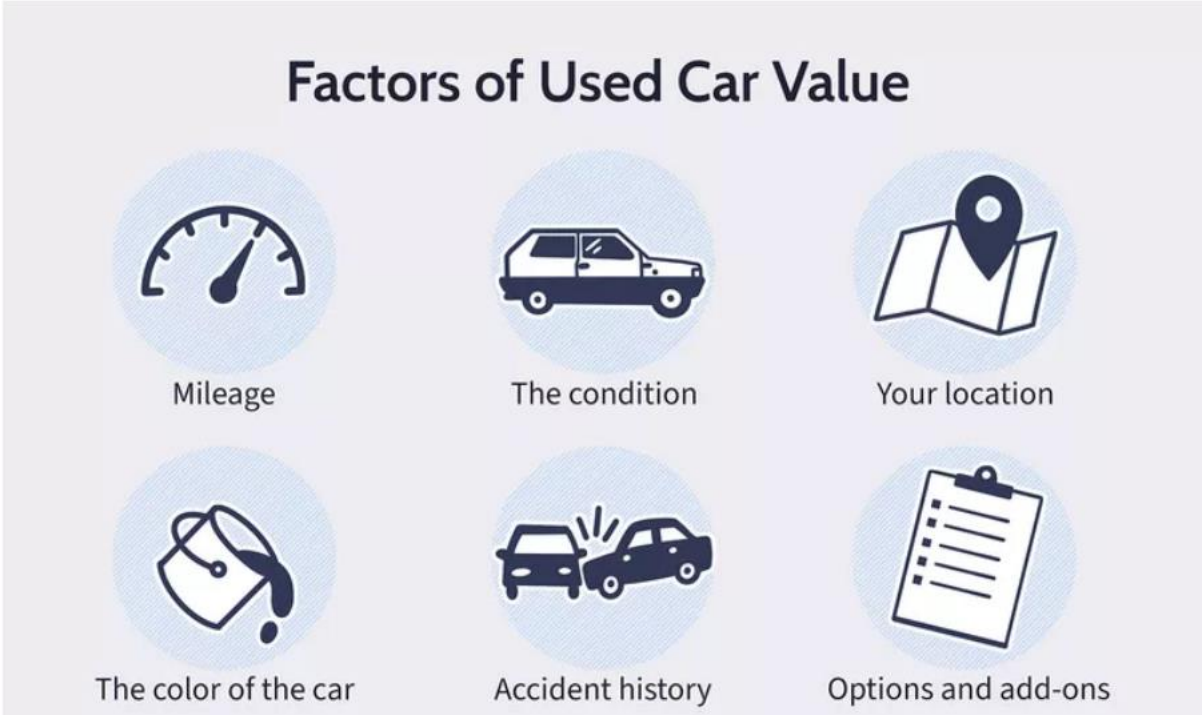
```
9.466622359363875
```

```
linearRegression.coef_
```

```
array([-3.2024792 , -0.53654814, -0.56352308,  1.26979282,  4.37603214,
        0.05842877, -0.35177362, -1.39416196, -0.44553453, -0.63116474,
       -0.31015978, -1.9827897 , -3.11907444, -3.64565231, -0.31808928,
       -0.15811326,  0.43451498, -0.46620612,  0.90592791,  1.30148724,
```

10.Final Product Prototype

Creating a detailed final product prototype of a used car price prediction AI model involves several components, including data preparation, model training, user interface development, and deployment. Below is a step-by-step breakdown of the prototype development process:



Data Collection and Preparation:

Gather a diverse and comprehensive dataset of historical used car sales data. The dataset should include car features (e.g., make, model, year, mileage, condition) and corresponding sale prices.

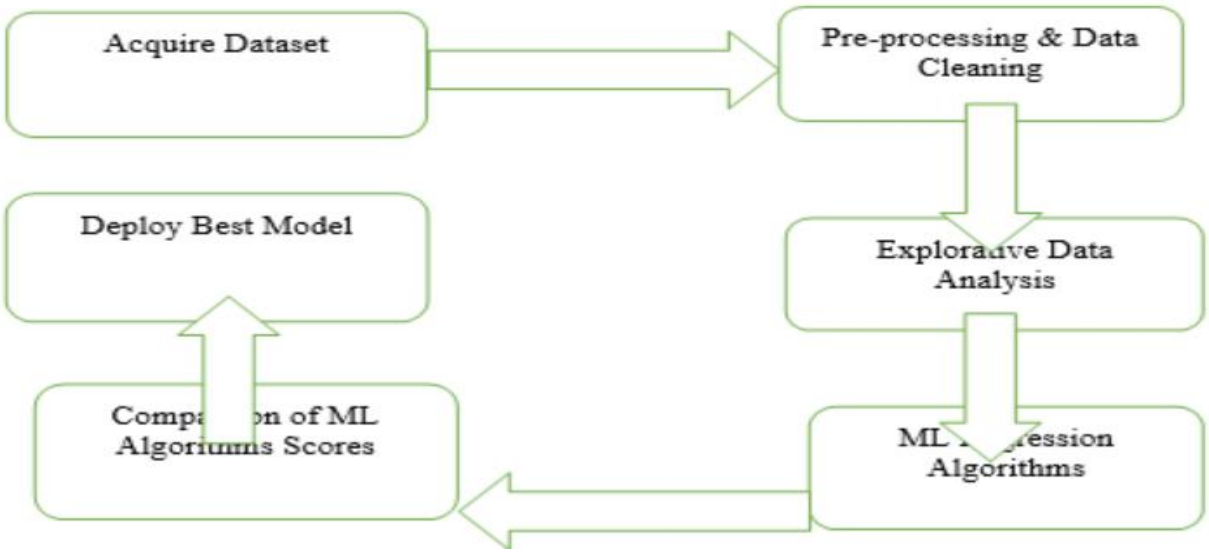
Perform data cleaning and preprocessing to handle missing values, outliers, and ensure data consistency. Split the dataset into training and testing sets for model evaluation.

Feature Engineering: Identify relevant features that influence used car prices, such as brand popularity, model year, mileage, condition, and location. Convert categorical features into numerical representations using techniques like one-hot encoding or label encoding.

Model Selection and Training: Choose an appropriate machine learning algorithm for regression tasks, such as linear regression, decision tree regression, random forests, or gradient boosting.

Train the selected model on the training dataset using the cleaned and engineered features.

Model Evaluation: Evaluate the trained model's performance on the testing dataset using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to assess its accuracy.



Backend Development: Create a backend server using Python or a similar language to handle user requests and process input data. Integrate the trained machine learning model into the backend to generate predictions based on user inputs.

Real-Time Predictions: Enable real-time predictions, providing users with instant estimated prices based on their input data.

Error Handling and User Feedback: Implement error handling to manage exceptions and provide meaningful feedback to users in case of invalid inputs or errors. Offer clear explanations for the factors influencing the predicted price to enhance transparency and user trust.

Privacy and Security: Ensure data privacy and implement measures to handle user data securely and anonymously.

Deployment: Host the web application on a cloud-based platform for accessibility and scalability.Test the application thoroughly to ensure performance and functionality.

Conclusion:

In conclusion, the car prediction AI model presents a promising opportunity in the automotive industry. By leveraging machine learning algorithms and historical car sales data, the AI model can accurately estimate the fair market value of used cars. This provides valuable insights to car sellers, buyers, dealerships, and insurance companies, enhancing decision-making processes and improving overall customer experiences.

The AI model's potential to optimize pricing strategies for car sellers, minimize financial risks, and provide fair compensation during insurance claims makes it a valuable tool for various stakeholders. Its transparent and data-driven approach fosters trust and credibility in the automotive market, leading to increased customer satisfaction and loyalty.

However, successful implementation and deployment of the car prediction AI model require careful considerations of data privacy, ethical concerns, and regulatory compliance. Ensuring the model's interpretability and fairness is essential to avoid biased predictions and discriminatory outcomes.

Overall, the car prediction AI model offers a competitive advantage and business opportunities for those willing to leverage advanced technology in the automotive industry. With continuous improvement, user feedback, and adherence to industry regulations, the AI model can be developed into a market-ready solution that brings value to the automotive ecosystem and helps shape the future of the used car market.