**Used Car Price Prediction- Review**

Sandip Badgujar 1 , Suyash Dingankar 2 , Sahil Landge 3 , Siddhesh Mote 4

1,2,3,4 Student, Department of Computer Engineering

M.H. Saboo Siddik Polytechnic

Byculla, Mumbai

**Guide:** Prof. Kausar Akumalla Ayub

Assistant Professor, Department of Computer Engineering

M.H. Saboo Siddik Polytechnic

Byculla, Mumbai

***Abstract---*: This review paper discusses the methods and techniques used to predict the prices of used cars using machine learning. The paper also examines the factors influencing car prices and the importance of data preprocessing. It highlights the importance of these techniques in helping buyers and sellers make informed decisions. The paper discusses different algorithms, the factors that influence car prices, and the significance of data preparation. The findings suggest that machine learning can significantly improve the accuracy of car price predictions.**

I. INTRODUCTION

Predicting vehicle prices accurately is important but challenging. The price of used cars is influenced by many factors such as age, mileage, brand, and condition. As the demand for used cars increases, accurately predicting their prices becomes essential for buyers and sellers. This paper aims to review the various machine learning techniques used in predicting car prices and to understand how these models can help in making informed decisions.

The demand for private cars is increasing globally, which is boosting the second-hand car market and creating business opportunities for both buyers and sellers. In many countries, buying a used car is a smart choice because they are usually more affordable. After using a car for a few years, owners can often sell it for a profit. However, several factors affect the price of a used car, such as its age and condition. Since used car prices vary, a model is needed to help evaluate their worth.

In this paper, we compared three methods—multiple linear regression, random forest regression, and gradient boosted regression trees—to create a price prediction model for used cars. We used data collected from e-commerce websites for our analysis. The main goal of this study is to find the best model for predicting used car prices.

II. LITERATURE REVIEW

Several studies have explored the use of machine learning for car price prediction. Many studies have looked into how machine learning can be used for car price prediction. Some researchers have found that methods like linear regression and decision trees can provide good results.The Decision Tree method creates a model that predicts the price based on decision rules derived from the data. It is easy to interpret and can handle non-linear relationships effectively.

The research paper highlights several studies that focus on predicting used car prices using various machine learning techniques. Many researchers have recognized the growing demand in the second-hand car market and the need for reliable price predictions. One notable study compared different machine learning models, revealing that gradient boosted regression trees achieved the highest accuracy, while traditional methods like linear regression had lower performance. This indicates that more advanced techniques can provide better insights into car pricing.

III. Methodology

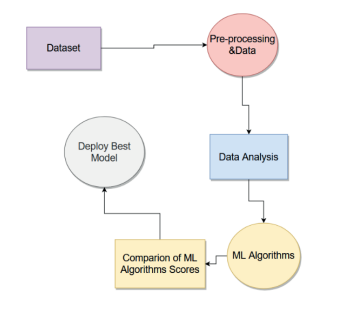
1. **Dataset**:
   * Utilize a dataset containing sale prices of various makes and models.
   * Include features such as car age, make and model, mileage, horsepower, and fuel economy.
2. **Machine Learning Algorithms**:
   * **Regression Techniques**: Focus on algorithms that provide continuous output values.
     + **Linear Regression**: A basic approach to model the relationship between features and price.
     + **Ridge Regression**: A variant of linear regression that includes regularization to prevent overfitting.
     + **Decision Tree Regressor**: A non-linear model that captures complex relationships between features and target prices.
3. **Performance Evaluation**:
   * Compare the performance of different algorithms using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared values.
   * Assess which algorithm provides the best predictions based on the dataset.

* **Key Factors Influencing Price**
* **Core Features**:
  + Age of the car
  + Make and model
  + Mileage
  + Horsepower
  + Fuel economy
* **Additional Factors**:
  + Type of fuel used
  + Interior style and safety features
  + Physical condition and ownership history
  + Market trends (e.g., fuel prices, consumer preferences)
* **User Interface**
* Develop a user-friendly interface that:
  + Accepts user input for car features.
  + Displays predicted pricing based on the selected features.
* **Challenges**
* **Data Availability**: Difficulty in obtaining comprehensive data on all influencing factors.
* **Complexity of Price Determination**: Many variables affect the resale value, making predictions challenging.

1. DATA AND METHOD

The methodology employed in this study for predicting car prices involves a multi-faceted approach, as illustrated in the conceptual framework (see Fig.1)

Figure 1. Conceptual Framework for the Car Price Prediction Process



The attributes collected for each vehicle included a wide array of features deemed relevant for price prediction:

•Brand, Model, Car Condition, Fuel Type

•Year of Manufacturing, Power (in kilowatts), Transmission Type

• Mileage, Color, City, State, Number of Doors

• Boolean attributes indicating the presence or absence of specific features such as Four Wheel

Drive, Navigation, Leather Seats, Alarm System, Aluminum Rims, Digital Air Conditioning,

Parking Sensors, Xenon Lights, Remote Unlock, Electric Rear Mirrors, Heated Seats, Panorama

Roof, Cruise Control, ABS, ESP, ASR

• Price, expressed in TL (Turkish liras) for the dataset.

**A) Dataset Collection**

This section explains how the data for the used cars was gathered. The dataset was obtained from a website called Kaggle and is stored in a CSV format, which is a common way to organize data in tables. The dataset includes 14 different pieces of information, such as a serial number, car name, location, mileage, fuel type, engine size, transmission type, kilometers driven, power, new price, year, number of seats, owner type, and the price of the car.

**B) Data Preprocessing**

This is an important step in preparing the data for machine learning. It involves several tasks:

i) Removing Non-numerical Parts from Numerical Features:

In this step, any non-numerical words from features like mileage, engine, and power are removed to make the data cleaner and easier to work with. The process involves converting the data into a list, splitting that list based on certain characters, and then saving the cleaned data back into a structured format.

ii) Converting Categorical Values into Numerical Values:

Categorical values, such as car name, location, fuel type, transmission type, and owner type, need to be changed into numbers because machine learning algorithms work better with numerical data. This conversion is done using a tool called Label Encoder, which is a Python package. The process includes selecting the categorical values and then changing them into numerical values.

iii) Separating the Target Variable:

In this step, we identify the specific value we want to predict, which in this case is the price of the car. The price is assigned to a variable called 'y', while all the other information (except for the price) is assigned to a variable called 'X'. This separation helps in organizing the data for the prediction model.

IV. RESULTS

Various models have been tested, with linear regression and decision trees being the most commonly used. The results indicate that these models can reduce prediction errors significantly compared to traditional methods.

i) Evaluation Dataset:  
We gathered our data from various shops in Bangladesh and from social media. After putting together our dataset, we tested the algorithms by dividing the data into two parts. We used 70% of the data for training the model and 30% for testing it.

A summary of the dataset is shown in Table II.

|  |  |
| --- | --- |
| **Attributes** | **Number of Count** |
| Total Data Collected | 50 |
| Data Used for Training | 35 |
| Data Used for Testing | 15 |

ii) Evaluation Measurement:  
To check how well our predictions worked, we looked at several factors, including precision-recall, execution time, and the accuracy of the algorithms. These measures help us understand how effective our model is.

V. DISCUSSION

The discussion highlights the challenges faced in car price prediction, such as data availability and the complexity of factors affecting prices. The Data Quality accuracy of predictions heavily relies on the quality of the data used. The other challenge is market dynamics in that the automotive market is influenced by various external factors such as economic conditions, consumer preferences.

After running the Machine Learning Algorithm, the next step is to assess how well the model performs using different performance metrics. Different metrics are used depending on the type of Machine Learning algorithm. For classification tasks, we typically look at metrics like Accuracy, Cross Validation, Precision, Recall, and F1 Score.

In our case, since we are predicting something (like car popularity), we use metrics such as Root Mean Square Error (RMSE) and Mean Square Error (MSE). However, because we don't have the actual output data, we can't directly measure how well our Machine Learning algorithms performed.

VI. CONCLUSION

In conclusion, machine learning techniques provide a promising approach to predicting used car prices. Despite the challenges, these models can help consumers make better purchasing decisions. The review highlights the effectiveness of various algorithms, particularly ensemble methods and neural networks, in capturing the complexities of car pricing. Future research should focus on refining these models, enhancing data quality, and exploring real-time prediction capabilities to adapt to the ever-changing automotive market.

VII. Future Scope

In the coming years, Machine Learning will greatly improve how we predict automobile prices and will become much more efficient. Deep Machine Learning and Neural Networks, which are types of Machine Learning, can analyze complex data, including past sales, economic factors, and new technologies. These models are constantly improving and can find small patterns that regular methods might miss.

Using real data, like economic trends and customer opinions from social media, can enhance prediction accuracy. This flexibility helps Machine Learning-based car price prediction systems adapt to fast-changing market needs.

In the future, we plan to collect more features and expand our dataset. We will use better techniques to improve accuracy and try advanced methods like Fuzzy Logic, Decision Trees, Artificial Neural Networks, and Ordinary Least Squares Regression (OLSR). We can also enhance this work by selecting more features for classification.

VIII. REFERENCES

[1] Nitis Monburinon, Suwat Rungpheung, Prajak Chertchom , Sabir Buya, Thongchai Kaewkiriya, Pitchayakit Boonpou’s “Prediction of Prices for Used Cars by Using Regression Models” IEEE, 2018.

[2] Snehit Shaprapawad, Premkumar Borugadda, Nirmala Koshiga’s “Car Price Prediction: An Application of Machine Learning” IEEE, June 2023.

[3] Rajiv Dahal, Rabin Timalsina’s “Used Car Price Prediction Using Linear Regression” IEEE, November 2023.

[4] Doan Van Thai, Luong Ngoc Son, Pham Vu Tien , Nguyen Nhat Anh, Nguyen Thi Ngoc Anh’s “Prediction Car Prices Using Quantify Qualitative Data and Knowledge-Based System” IEEE, 2019.

[5] Ning Sun, Hongxi Bai, Yuxia Geng, Huizhu Shi’s “Price Evaluation Model in Second-hand Car System Based on BP Neural Network Theory” IEEE, 2017.

[6] Adriana Chis, Jarmo Lundén, Visa Koivunen’s “Reinforcement Learning-Based Plug-in Electric Vehicle Charging with Forecasted Price” IEEE, 2016.

[7] Sajjad Mozaffari, Omar Y. Al-Jarrah, Mehrdad Dianati, Paul Jennings, Alexandros Mouzakitis’s “Deep Learning-based Vehicle Behaviour Prediction for Autonomous Driving Applications” IEEE, 2020

[8] Sunakshi Mamgain, Swati Vipsita, Srikant Kumar, Kabita Manjari Nayak’s “Car Popularity Prediction: A Machine Learning Approach” IEEE, 2018.

[9] Anamika Das Mou, Protap Kumar Saha, Sumiya Akter Nisher, Anirban Saha’s “A Comprehensive Study of Machine Learning Algorithms for Predicting Car Purchase Based on Customers' Demands” IEEE, 2021.

[10] Nur Oktavin Idris, Aspian Achban, Siti Andini Utiarahman, Jorry Karim, Fuad Pontoiyo’s “Predicting the Selling Price of Cars Using Business Intelligence with the Feed-forward Backpropagation Algorithms.” IEEE, 2020.

[11] Nabarun Pal, Priya Arora, Dhanasekar Sundararaman, Puneet Kohli, Sai Sumanth Palakurthy’s “How Much Is My Car Worth? A Methodology for Predicting Used Cars Prices Using Random Forest” Springer, 2018.

[12] David Von Dollen, Florian Neukart, Daniel Weimer, Thomas Back’s “Predicting Vehicle Prices via Quantum-Assisted Feature Selection” Springer, 2023.

[13] Zhenchang Xia, Shan Xue, Libing Wu, Jiaxin Sun, Yanjiao Chen, Rui Zhang’s “Passenger Car Sales Prediction Based on XGBoost” Springer, 2020.

[14] Ahmed Fathalla, Ahmad Salah, Kenli Li, Keqin Li, Francesco Piccialli’s “Deep End-to-End Learning for Price Prediction of Second-Hand Items” Springer, 2020.

[15] Naser Shanti, Akram Assi, Hamza Shakhshir, Adnan Salman’s “Machine Learning-Powered Mobile App for Predicting Used Car Prices” ACM, 2021.

[16] Kiran Madhusudhanan, Gunnar Behrens, Maximilian Stubbemann, Lars Schmidt-Thieme’s “Probabilistic Tabular Regression for Used Car Pricing” ACM, 2024.

[17] Subba Rao Peram, Mundru Sri Lekha, Kothagundla Sai Sri Jyothsna, Kunisetty Lakshmi Pallavi, V. Hema’s “Vehicle Count Prediction Using Machine Learning” ELSEVIER, 2022.

[18] Giovanni Carvalho, Darniton Viana, Luciano Barbosa, Tsang Ing Ren’s “Combining Structured and Unstructured Data using” ELSEVIER, 2023.

[19] Chao Ma, Zhenbing Liu, Zhiguang Cao, Wen Song, Jie Zhang, Weiliang Zeng’s “Cost-sensitive Deep Forest for Price Prediction” ELSEVIER, 2020.

[20] Stefan Lessmann, Stefan Voß’s “Car Resale Price Forecasting: The Impact of Regression Method, Private Information, and Heterogeneity on Forecast Accuracy” ELSEVIER, 2017.

[21] Filip Mandys’s “Search for an Affordable Clean Car: Pricing of Conventional and Clean Automobiles” ELSEVIER, 2023.

[22] Dean Fantazzini, Zhamal Toktamysova’s “Forecasting German Car Sales Using Google Data and Multivariate Models” ELSEVIER,2015.

[23] Lakshmi Prasad Mudarakola, D Shabda Prakash, K L N Shashidhar, D Yaswanth’s “Car Price Prediction Using Machine Learning” INTERNATIONAL REASEARCH JOURNAL OF MODERNIZATION IN ENGINEERING TECHNOLOGY AND SCIENCE, 2024.

[24] Manuparna Nandan, Debolina Ghosh’s “Pre-owned Car Price Prediction by Employing Machine Learning Techniques” INTERNATIONAL JOURNAL FOR MULTIDISCIPLINARY RESEARCH, 2023.

[25] Pranshu Mishra, Rishabh Salgarkar, Anushri Nalavade, Arpit Kumar, Sudhanshu Gonge, Rahul Joshi, Kailash Shaw, Preeti Mulay, Ketan Kotecha’s “Predicting the Price of a Vehicle using Machine Learning” ELSEVIER, 2023.

[26] Abhishek Pandey, Vanshika Rastogi, Sanika Singh’s “Car’s Selling Price Prediction using Random Forest Machine Learning Algorithm” ELSEVIER, 2020.

[27] Mr. B.V. Ramana, Alla Govardhini, Ande Divya Tejaswini, Bora Yuvasai Krishna Reddy, N.V Aditya, Annadata Aravind’s “Car Price Prediction Using Random Forest Algorithm” ELSEVIER, 2024.

[28] Snehil Gupta, Dr. Sowmya K S’s “Car Price Prediction Using Machine Learning” INTERNATIONAL REASEARCH JOURNAL OF MODERNIZATION IN ENGINEERING TECHNOLOGY AND SCIENCE, 2024.

[29] Ashutosh Datt Sharma, Vibhor Sharma’s “Used Car Price Prediction Using Linear Regression Model” INTERNATIONAL REASEARCH JOURNAL OF MODERNIZATION IN ENGINEERING TECHNOLOGY AND SCIENCE, 2020.

[30] Prof. Dipti A. Gaikwad, Pratik S. Suwarnakar, Yash R. Mahajan, Amita U. Petkar, Shreyasi G. Theurkar’s “Used Car Price Prediction Using Random Forest Algorithm” INTERNATIONAL JOURNAL FOR MULTIDISCIPLINARY RESEARCH, 2023.