

Car Price Predictor: Unlocking Insights for Used Car Buyers and Seller

S. Aravind

Hindustan Institute of Technology and Science
Chennai-India
22247010@student.hindustanuniv.ac.in

Dr. A Rama Prasath

Associate Professor
Hindustan Institute of Technology and Science
Chennai-India
rprasath@hindustanuniv.ac.in

Abstract

The growing market for old cars is causing an evolutionary shift in the automotive sector. The difficult part of the purchasing and selling method for old cars is the rating of these vehicles because buyers look into more affordable and environmentally friendly modes of transportation. This paper offers a thorough analysis of the use of machine learning techniques to estimate used automobile prices. Several factors including mileage and other relevant characteristics which impact the nature of used vehicle costs are gathered and examined as part of our methodology. Our machine learning algorithms seek to understand the complex correlations among these qualities and used car sale prices. It is done by utilising an extensive database that includes an extensive variety of cars. To improve the capacity of the model for detecting minute trends in the data, methods such as feature engineering are used. Pre-processing the dataset properly ensures that the prediction models are stable by handling values that are lacking, anomalies, and variables that are categorical. The machine learning-powered predictive algorithms that have been built are useful resources for customers as well as sellers in the marketplace for used cars. Our strategy enhances accountability, effectiveness, and sound choice-making in the ever-changing used car market by offering precise and data-driven automobile pricing calculations.

Keywords— Artificial intelligence, Machine learning, Forecasting system, Data analytics, Prediction system.

I. INTRODUCTION

The car industry is changing dramatically as consumers choose more affordable and environmentally friendly modes of transportation. As more and more people look to the used automobile market for affordable alternatives, the appraisal of used automobiles becomes crucial to the entire buying or selling processes. Understanding how important it is to price correctly in this ever-changing marketplace, this study sets out to investigate how to forecast used car prices by effectively utilising cutting-edge machine learning technologies. The market for used automobiles is rising in this period of changing customer tastes, driven by factors including cost-effectiveness, long-term viability and a wide selection of car options. As a result, knowing and estimating the market worth of used cars remains crucial for negotiating this unstable environment for consumers as well as sellers. In order to create forecasting algorithms which can offer deep understanding regarding the changing trends of used automobile costs, this research explores the complexities of machine learning techniques. The analysis procedure includes a thorough examination of numerous factors. It also concentrates on mileage and various significant characteristics that are crucial in determining the price environment. Our machine learning algorithms are intended to understand the complicated links that are present among these attributes as well as the market values of used automobiles. This is done by utilising an extensive database that covers a wide range of cars. The capacity of the model is improved to identify minute variations which are hidden in the available data by utilising sophisticated feature engineering methods. This approach fits to the changing demands of customers and other interested parties by providing precise and statistical predictions of previously owned vehicle costs. Hence it promotes a more knowledgeable and effective utilised car trading system.

This project's main goal is to apply machine learning techniques for accurately predicting used automobile costs. In order to do this, strong models must be created, which includes machine learning methods like decision trees and linear regression. It also needs extensive evaluations of important factors like the type of fuel, mileage, and branding. Modern feature engineering methods are going to be utilised for obtaining hidden trends from the information set. In addition to that, rigorous data collection guarantees the models' dependability. By giving honest and data-driven predictions, the project hopes to offer useful applications to the consumers and sellers in the resale automobile market. Prediction accuracy will be measured through assessment of models using important indicators like mean absolute error, which in turn will improve effectiveness and honesty in used automobile deals. In order to address changing industry trends and issues, the project will also investigate possible future developments. Eventually, the goal is to provide significant knowledge into the area of machine learning because it relates to used automobile cost.

The second part of the paper consists of background research investigation, while the third section discusses system issues and system design techniques. The fourth part examines the structure of the system as well as execution methods. The collected findings and remarks are contained in the remaining sections of the paper.

II. LITERATURE REVIEW

Bambang Kriswantara et al. (2022) This paper concentrate on the decrease in the buying power of individuals. This is because the worldwide outbreak has halted all operations, especially the drop in the global economy. A previously owned car can be a person's second choice if they wish to purchase a vehicle. Calculating costs from past customer transactions is the primary responsibility of certain used automobile sales dealerships. The cost estimation procedure requires considerable time because many elements of the car, such as the type of fuel, miles driven, and many more need to be focused. This in turn impacts the factors that influence car costs. Estimating the cost of used cars can be achieved by using a model that is capable of acquiring knowledge from past customer transactions, which is one of the functions of machine learning.

K.Samruddhi et al. (2020) Calculating the cost of used cars is one of the important and fascinating areas of investigation. The market for used cars has seen a rise in demand, which has boosted business for consumers as well as sellers. Since the cost of automobiles fluctuates based on numerous essential elements, precise and trustworthy projection demands specialist expertise in the sector. In order to evaluate used automobile prices, this article presented a model using supervised machine learning based on the KNN (K Nearest Neighbour) regression technique. The data of second-hand automobiles from the Kaggle website is gathered to train this model. This experiment assessed the data using various trained and tested ratios. Consequently, the suggested model's accuracy is approximately 85%, and it is fitted as the best one.

Milunovich et al. (2023) In the presented paper Fifteen machine learning strategies are trained utilising an overall 141 automotive features. The model's scores are provided based on two key performance indicators regarding the extent to which they can forecast used car pricing. Furthermore, the model's evaluation procedure is executed using the model confidence set (MCS) method. When the examination is performed on the in-sample (training sample) basis, the LightGBM algorithm is the only model in the MCS that scores better than the other fourteen algorithms. This is based on both RMSE and MAE parameters. In an independent test dataset, LightGBM places third by RMSE but remains top in MAE when the examination is conducted out-of-sample.

Pattabiraman Venkatasubbu et al. (2020) The development of internet-based platforms has made it easier for customers as well as sellers to learn more about the developments and patterns that influence a used car's value on the marketplace. With the help of past customer data and a predetermined set of elements, a statistical model is created that can forecast the price of a used car. This is done by employing machine learning algorithms like Lasso Regression, Multiple Regression, and Regression Trees. Additionally, the best model is chosen by evaluating the predictive power of each one.

Liu et al. (2022) A used automobile estimation system that utilises PSO-GRA-BPNN was suggested. The particle swarm optimisation approach was incorporated with the conventional BP neural network to improve it. The findings indicate that a specific linear connection is only presented between the cost of new cars, engine performance, and the used cars, with the correlation coefficient of each being more than 0.6. The result of other factors on second-hand vehicle cost is minimal. Additionally, the connection among the cost of new cars and the cost of second-hand vehicles, as well as the movement, kilometres travelled, gearshift type, usage of fuel, and the registration time, is larger than 0.7. The MAPE of the PSO-GRA-BPNN model put forward in this paper is 3.936%, meaning it is 30.041% lesser than the error rates of the remaining three alternatives. This is obtained by

comparing with the conventional BPNN model along with the various linear regression, random forest, and support vector machine regression methods provided by different investigators.

C. Jin et al. (2021) The goal of this work is to develop a system which will forecast the fair market value of used automobiles according to a number of factors, such as distance travelled, production year, efficiency of fuel, gearbox, vehicle tax, type of fuel, and the size of the engine. In the second-hand vehicle industry, this model can help manufacturers, consumers, and vendors of automobiles. Once it's done, using the data that people enter, it may produce almost correct prices of cars. The fields of machine learning and data science are included in the model's development process. The data set for used vehicles was taken from second hand vehicle listings through skimming.

III. PROBLEM DEFINITION

- The approach ignores important contributing elements by focusing just on a small number of indicators, like years and mileage. The value of an automobile is only partially represented as a result of this restricted area.
- The current system limits capacity for learning and is unable to capture the complicated connections seen in a variety of dataset. This is because of its absence of accuracy and ability to forecast compared to more sophisticated machine learning models.
- Its capacity to adapt and increase precision as time passes is constrained. This is because of its inability to frequently improve and update its value estimations in accordance with latest facts.

IV. METHODOLOGY

A.

System architecture

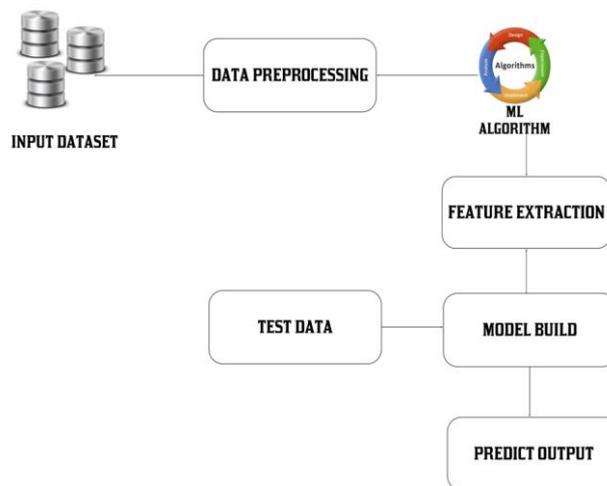


Fig 7. System architecture of proposed prediction model

Fig 7. Shows the system architecture of proposed used car prediction model. The suggested used automobile valuation system takes an innovative approach by utilising cutting-edge machine learning techniques. Moving from the restrictions of the current system, this creative method uses a wide range of factors. These include model performance, and so on to achieve an outstanding level of accuracy in used automobile valuation. By employing the use of feature engineering and thorough data preparation, the algorithm guarantees an accurate understanding of difficult connections across different sets of data. Its ability to adjust to changing market conditions and providing current reviews by constant learning from fresh data sets is essential to its success. The suggested system's technology, flexibility, accountability, and capacity to provide statistical data represent an exciting development in the used automobile market. This is by supplying customers with precise and knowledgeable choice-making instruments.

- Through the integration of an extensive range of criteria, such as current market conditions and vehicle requirements, the suggested approach attains a remarkable degree of precision in used automobile valuation.
- Strong data preparation and smart feature engineering approaches improve the ability of the system for recognising complex connections across a variety of information sets.

• The suggested system stands out since it has great flexibility in responding to changing market conditions. The system's ability to continuously learn from new data patterns enables it to deliver updated costs, guaranteeing that customers get correct and dynamic valuations in a constantly changing market.

A critical stage in the used car pricing prediction project is the data preparation component, which converts unprocessed data into a form that's suitable for machine learning algorithm training. This part includes a number of procedures to guarantee the importance, regularity, and overall integrity of the database. To keep the dataset intact, missing values are first handled either by restoration or elimination and then outliers have to be dealt accordingly.

A thorough evaluation and application of regression methods constitute the fundamental approach in the machine learning algorithm component of the utilised vehicle cost estimation study. The investigation includes algorithms like Decision Trees and Linear Regression, with an emphasis on combinations of methods to improve accuracy in prediction. The process of hyper parameter tuning involves the utilisation of methods such as grid-based searching and random searching to identify the best parameters for every method. The methods' scores are used to analyse the significance of the characteristics, which contributes to identifying the factors that have a major impact.

The best model is chosen using the comparison study as a reference. To bring together results from multiple models, experiments are conducted with ensemble modelling approaches such as stacking or blending. For successful implementation, developed models are lined up, and clarity and clearness are taken into account. If necessary, techniques for dealing with inaccurate information are put into practice. With respect to the given variables, the application is intended to provide flexible, readable, and precise models for utilised vehicle price estimation.

Performance evaluation

The performance of the system is evaluated through accuracy, precision, recall and F1score measurement.

The accuracy of the prediction model is calculated based on number of positive outcomes of the system arrived from the total number of predictions.

$$\text{Accuracy} = \frac{TP+TN}{(TP+TN+FP+FN)}$$

Precision is the evaluation metric detect the amount of true positive predictions found from the total predictions. Precision is calculated from the following formula.

$$\text{Precision} = \frac{TP}{TP+FP}$$

The recall is the evaluation metric helpful to predict the correctly classified true values from the real true value expected.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1score detects the effectiveness of the proposed model by evaluating the relationship between Precision, and Recall.

$$\text{F1score} = \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$$

V. DISCUSSIONS

Keeping various challenges in the existing implementations on used car prediction system, the lack of accuracy on the system need to be focused. The presented approach considers ensemble approach. The existing linear regression model, decision tree model, enhance the accuracy of the system.

Further the presented system is developed using python software. The integrated python environment is simulated through Google collab. The results are comparatively validated through accuracy, precision, recall and F1score for the proposed approach, compared with existing state of art approaches.

VI. CONCLUSION

The becoming business sector for old vehicles is causing a transformative change in the automotive area. The troublesome aspect of the buying and selling technique for old vehicles is the rating of these vehicles since purchasers investigate more reasonable and harmless to the ecosystem methods of transportation. This paper offers a careful investigation of the utilization of machine learning methods to gauge utilized car costs. A few variables including mileage and other important qualities which influence the idea of pre-owned car costs are assembled and inspected as a component of our strategy. Our machine learning algorithms try to comprehend the mind boggling relationships among these characteristics and pre-owned vehicle deal costs. It is finished by using a broad data set that incorporates a broad assortment of vehicles. To work on the limit of the model for identifying minute patterns in the information, techniques, for example, highlight designing are utilized. Pre-handling the dataset appropriately guarantees that the forecast models are steady by dealing with values that are deficient with regards to, irregularities, and factors that are straight out. The machine learning-controlled prescient algorithms that have been fabricated are helpful assets for clients as well as dealers in the commercial

center for utilized vehicles. Our technique improves responsibility, adequacy, and sound decision making in the steadily changing pre-owned vehicle market by offering exact and information driven car valuing estimations.

REFERENCES

- [1] C. Flores, P. Merdrignac, R. de Charette, F. Navas, V. Milanés and F. Nashashibi, "A Cooperative Car-Following/Emergency Braking System With Prediction-Based Pedestrian Avoidance Capabilities," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1837-1846, May 2019, doi: 10.1109/TITS.2018.2841644.
- [2] S. Shaprapawad, P. Borugadda and N. Koshika, "Car Price Prediction:An Application of Machine Learning," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 242-248, doi: 10.1109/ICICT57646.2023.10134142.
- [3] Noor, Kanwal, and Sadaqat Jan. "Vehicle price prediction system using machine learning techniques." *International Journal of Computer Applications* 167.9 (2017): 27-31.
- [4] Gegic, Enis, et al. "Car price prediction using machine learning techniques." *TEM Journal* 8.1 (2019): 113.
- [5] Pudaruth, Sameerchand. "Predicting the price of used cars using machine learning techniques." *Int. J. Inf. Comput. Technol* 4.7 (2014): 753-764.
- [6] Gajera, P., Gondaliya, A., & Kavathiya, J. (2021). Old car price prediction with machine learning. *Int. Res. J. Mod. Eng. Technol. Sci*, 3, 284-290.
- [7] Bambang Kriswantara , Rifki Sadikin , "Used Car Price Prediction with Random Forest Regressor Model" 2022..
- [8] K.Samruddhi, Dr. R.Ashok Kumar, "Used Car Price Prediction using K-Nearest Neighbor Based Model" 2020.
- [9] Milunovich, George and Wu, Lijun and Zhao, Yuanyuan, Forecasting the Prices of Used Cars: A Comparative Analysis of Supervised Learning Algorithms (January 23, 2023). Available at SSRN: <https://ssrn.com/abstract=4334750> or <http://dx.doi.org/10.2139/ssrn.4334750>
- [10] Pattabiraman Venkatasubbu, Mukkesh Ganesh," Used Cars Price Prediction using Supervised Learning Techniques" ,2020
- [11] Liu, E.; Li, J.; Zheng, A.; Liu, H.; Jiang, T. Research on the Prediction Model of the Used Car Price in View of the PSO-GRA-BP Neural Network. *Sustainability* 2022, 14, 8993. <https://doi.org/10.3390/su14158993>
- [12] C. Jin, "Price Prediction of Used Cars Using Machine Learning," 2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT), Chongqing, China, 2021, pp. 223-230, doi: 10.1109/ICESIT53460.2021.9696839.