



Used Vehicle Price Prediction with Machine Learning

Mrs. Anuja S. Phapale^{1*}, Mohammed Munawwar Rangila²
Department of Information Technology, AISSMS Institute of Information
Technology, Pune, Maharashtra, India.

Corresponding Author: Munawwar rangilamohammedmunawwar@gmail.com

ARTICLE INFO

Keywords: Prediction,
Machine Learning,
Regression, Analyze

Received : 10, August *Revised :*
20, September *Accepted:* 29,
October

©2023 Phapale, Rangila: This
is an open-access article
distributed under the terms
of the [Creative Commons
Attribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

The automotive industry has experienced significant growth and diversification recently years. With this growth comes an increase in the demand for used vehicles. However, determining the accurate price of a used vehicle is a complex task due to various factors such as age, mileage, condition, and market trends. Efficient and accurate price prediction and classification are essential for both buyers and sellers in order to make informed decisions. To tackle this challenge, supervised machine learning techniques have gained prominence in recent years. These historical leverage techniques data to train models that can predict and classify the prices of used vehicles. By analyzing large datasets of past sales, these models can identify patterns and relationships between different features of a vehicle and its market value.

INTRODUCTION

The automotive industry has experienced significant growth and diversification in recent years. With this growth comes an increase in the demand for used vehicles. However, determining the accurate price of a used vehicle is a complex task due to various factors such as age, mileage, condition, and market trends.

Efficient and accurate price prediction and classification are essential for both buyers and sellers in order to make informed decisions. To tackle this challenge, supervised machine learning techniques have gained prominence in recent years. These techniques leverage historical data to train models that can predict and classify the prices of used vehicles. By analyzing large datasets of past sales, these models can identify patterns and relationships between different features of a vehicle and its market value.

The learned models can then be utilized to predict accurate prices for similar vehicles in the future. The significance of accurate price prediction and classification of used vehicles goes beyond individual buyers and sellers. It can also aid in decision-making for insurance companies, vehicle manufacturers, and lending institutions. Moreover, these models can contribute to improving the overall transparency and fairness in the used vehicle market.

In this research paper, we aim to explore the application of supervised machine learning algorithms in predicting and classifying the prices of used vehicles. We will analyze a comprehensive dataset of used vehicle listings, considering various features such as make, model, year, mileage, condition, and geographical location. By comparing different algorithms and evaluating their performance, we hope to contribute to the development of more effective and reliable methods for price prediction and classification in the used vehicle market.

LITERATURE REVIEW

The automotive industry has witnessed remarkable growth and diversification in recent years, leading to an increased demand for used vehicles. However, accurately determining the price of a used vehicle has become a complex task due to various factors, including age, mileage, condition, and market trends. Efficient and accurate price prediction and classification are crucial for both buyers and sellers to make informed decisions in this dynamic market. In response to this challenge, supervised machine learning techniques have gained prominence in recent years.

Supervised machine learning techniques leverage historical data to train models capable of predicting and classifying the prices of used vehicles. By analyzing vast datasets of past sales, these models identify patterns and relationships between different features of a vehicle and its market value.

The acquired knowledge enables the models to predict accurate prices for similar vehicles in the future. This application of machine learning holds significant potential for improving decision-making processes not only for

individual buyers and sellers but also for insurance companies, vehicle manufacturers, and lending institutions.

The accurate prediction and classification of used vehicle prices carry implications beyond the immediate transaction. They also contribute to enhancing transparency and fairness in the used vehicle market. By employing supervised machine learning algorithms, we aim to explore their application in predicting and classifying the prices of used vehicles. Our research will involve the thorough analysis of a comprehensive dataset of used vehicle listings, taking into account various features such as make, model, year, mileage, condition, and geographical location.

The primary focus of our research paper is to compare different machine learning algorithms and evaluate their performance in car price prediction and classification. By conducting in-depth analyses and assessing the effectiveness and reliability of these algorithms, we aim to contribute to the development of more sophisticated and dependable methods for price prediction and classification in the used vehicle market.

The potential impact of our research extends beyond individual buyers and sellers. Insurance companies can leverage accurate price prediction models to determine appropriate insurance premiums, while vehicle manufacturers can make informed decisions regarding pricing and marketing strategies. Similarly, lending institutions can assess the fair value of a used vehicle to determine loan amounts, interest rates, and repayment terms.

The anticipated outcomes of this research include the identification of the most effective machine learning algorithms for car price prediction and classification. By analyzing large datasets and considering various influencing factors, we hope to uncover new insights and patterns that will contribute to the accuracy and reliability of these prediction models. Furthermore, our research aims to lay the foundation for future studies in the field of car pricing, potentially leading to the refinement of existing models and the development of more advanced techniques.

In conclusion, the application of supervised machine learning techniques in predicting and classifying the prices of used vehicles is gaining prominence in the automotive industry. Our research aims to contribute to this field by analyzing a comprehensive dataset and evaluating different machine learning algorithms. The expected outcomes include the development of more effective and reliable methods for car price prediction and classification, ultimately benefiting individual buyers, sellers, insurance companies, vehicle manufacturers, and lending institutions.

METHODOLOGY

In this research paper, we propose a methodology for building a car price prediction system using various supervised learning algorithms. The goal is to develop a robust and accurate system that can predict the prices of

used vehicles based on their relevant features. The methodology comprises several steps, including data collection, preprocessing, feature selection, model training, evaluation, and performance comparison.

The first step in our methodology is data collection. We gather a comprehensive dataset of used vehicle listings from various sources, including online marketplaces, classified ads, and dealership databases. The dataset contains various attributes, such as make, model, year, mileage, condition, and geographical location.

Once the data is collected, the next step is data preprocessing. This involves cleaning the dataset by handling missing values, outliers, and inconsistent entries. We also normalize numerical attributes and encode categorical variables to ensure compatibility with the machine learning algorithms.

After preprocessing the data, the next step is feature selection. We employ feature selection techniques to identify the most relevant attributes that have a significant impact on car prices. This helps in reducing dimensionality and improving the efficiency of the prediction models. We consider methods such as correlation analysis, statistical tests, and information gain to select the optimal set of features.

With the selected features, we proceed to the model training phase. We experiment with different supervised learning algorithms, including linear regression, decision trees, random forests, support vector machines, and neural networks. We split the dataset into training and testing subsets to train the models on the training data and evaluate their performance on the testing data.

To evaluate the performance of the models, we employ various evaluation metrics such as mean squared error, mean absolute error, and R-squared value. These metrics help us quantify the accuracy and effectiveness of the prediction models. We also perform cross-validation to ensure that the models generalize well on unseen data and are not overfitting to the training set.

Finally, we compare the performance of the different supervised learning algorithms. We analyze the results obtained from each algorithm and identify the strengths and weaknesses of each approach. This allows us to determine which algorithm performs best for car price prediction and select the most suitable one for our system.

In summary, our methodology includes data collection, preprocessing, feature selection, model training, evaluation, and performance comparison. By following this methodology, we aim to develop a car price prediction system using various supervised learning algorithms. The research findings will contribute to

The advancement of the field and provide valuable insights into predicting and understanding car prices using machine learning techniques.

RESEARCH RESULT AND DISCUSSION

In our study on car price prediction utilizing a variety of supervised machine learning algorithms, we undertook a comprehensive data collection process from an online seller of used cars. We identified key features that strongly influence the price of these vehicles. To ensure the reliability of our analysis, we removed non-available values and excluded features that were not directly relevant to price prediction.

The primary phase of our research involved applying a range of supervised machine learning techniques to the first dataset. We then compared the validation results with the price prediction outputs from the second dataset, which was specifically curated to include essential features. Among the algorithms tested, the linear regression model demonstrated the highest accuracy when essential features such as price and model were available. This success was achieved across different vehicle types, considering their usage condition and associated prices.

Additionally, our study encompassed various techniques for numeric data pre-processing and text analysis, which were essential for handling unstructured data in the used car listings. This approach ensured that the predictions were based on comprehensive and well-processed information.

Our research revealed that the competitive advantage of second-hand market trend prediction, achieved through data mining and analysis, is notable. This advantage includes the ability to determine the optimal price for observed vehicles, thereby mitigating misclassification and associated risks. Moreover, it enhances customer awareness of market dynamics, leading to more accurate buying decisions.

As part of our future research direction, we plan to expand our analysis by incorporating additional datasets gathered over the next two quarters. These datasets will further enrich our understanding of the used vehicle market through the application of data mining and various machine learning techniques. Additionally, model validation will be carried out using different optimization methods. This research expansion will involve the inclusion of variables related to the vehicle's condition and drive type, factors that can significantly impact price prediction accuracy.

CONCLUSION

In conclusion, our research focused on utilizing supervised learning methods for car price prediction. We employed various machine learning algorithms such as linear regression, decision trees, random forests, support vector machines, k-nearest neighbors. The results obtained demonstrated the effectiveness of these techniques in accurately predicting car prices based on a given set of features. However, it is important to acknowledge the limitations of our study. These include the reliance on a limited dataset, which may not capture the full complexity of the car market, and the potential for overfitting our models to the training data. Despite these

limitations, our research provides valuable insights into the application of machine learning techniques for car price prediction, with potential applications in the automotive industry and consumer decision-making.

ADVANCED RESEARCH

This research still has limitations so further research needs to be done on this topic.

REFERENCES

- Bharambe, P.P.; Bagul, B.; Dandekar, S.; Ingle, P. Used Car Price Prediction using Different Machine Learning Algorithms. *Int. J. Res. Appl. Sci. Eng. Technol.* 2022, 10, 773–778.
- C K Puteri and L N Safitri, Citation C K Puteri and L N Safitri 2020 *J. Phys.: Conf. Ser.* 1469 012143 DOI 10.1088/1742-6596/1469/1/012143
- Puteri, C.K.; Safitri, L.N. Analysis of linear regression on used car sales in Indonesia. *J. Phys. Conf. Ser.* 2020, 1469, 012143
- Samruddhi, K.; Kumar, R.A. Used Car Price Prediction using K-Nearest Neighbor Based Model. *Int. J. Innov. Res. Appl. Sci. Eng.* 2020, 4, 629–632.
- Siva, R.; M, A. Linear Regression Algorithm Based Price Prediction of Car and Accuracy Comparison with Support Vector Machine Algorithm. *ECS Trans.* 2022, 107, 12953–12964.