www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

USED CAR PRICE PREDICTION USING MACHINE LEARNING

Mr.K.Sreenivasa Reddy [1]
Assistant Professor
Department of
CSE(DS)
TKR College of Engineering
and Technology
ksrinivasreddy@tkrcet.com

V.Vaishnavi [2]
B. Tech(Scholar)
Department of
CSE(DS)
TKR College of Engineering
and Technology
vaishnavireddyvelpoor@gmail.com

M.Fani Goud [3]
B. Tech(Scholar)
Department of
CSE(DS)
TKR College of Engineering
and Technology
mfanigoud@gmail.com

P.Abhiram [4]
B. Tech(Scholar)
Department of
CSE(DS)
TKR College of Engineering
and Technology
abhiram2709@gmail.com

K.Manigoud [5]
B. Tech(Scholar)
Department of
CSE(DS)
TKR College of Engineering
and Technology
manigoudkollu@gmail.com

ABSTRACT

The manufacturer sets the price of a new car in the industry, with the government incurring some additional expenditures in the form of taxes. Customers purchasing a new car may thus be confident that their investment will be worthwhile. However, due to rising new car prices and buyers' financial inability to purchase them, used car sales are increasing globally. A USED CAR PRICE PREDICTION SYSTEM that efficiently assesses the car's worthiness using a range of factors is required. The current system involves dealers deciding prices at random, leaving buyers unaware of the car's actual worth. Sellers, too, lack knowledge of appropriate pricing. To address this issue, a highly effective model is proposed. Regression algorithms are employed, producing continuous values rather than classified values. This allows for precise price estimation rather than generalized price ranges. A user interface has also been created to allow users to input data and view predictions.

KEYWORDS: Used Car Price Prediction, Regression Algorithms, Machine Learning, Linear Regression, Ridge and Lasso Regression, Bayesian Ridge Regression, Decision Tree, Random Forest, XGBoost, Gradient Boosting.

1.INTRODUCTION

The used car market has evolved significantly over the last decade, with buyers and sellers relying increasingly on online platforms and algorithms to determine car prices. Accurate pricing of used cars is critical for both buyers and

sellers as it ensures a fair exchange, prevents market inefficiencies, and builds consumer trust. Machine learning (ML) algorithms have demonstrated considerable potential in predicting the prices of used cars by leveraging historical data, which includes various factors such as make, model, year of manufacture, mileage, engine type, and additional features. With the growth of e-commerce platforms, there is a significant shift in the way used car pricing is handled, as traditional methods such as manual appraisal and dealer estimation are being replaced by data-driven approaches.

Machine learning techniques can automate the process of price prediction, making it faster, more accurate, and transparent. Several machine learning algorithms, including decision trees, support vector machines (SVM), and artificial neural networks (ANNs), have been employed to predict car prices with varying degrees of success. These models work by learning from historical data and extracting patterns that help predict future trends. The introduction of artificial intelligence (AI) in the used car market has opened new avenues for dynamic pricing, where prices are adjusted in real-time based on various factors such as demand, supply, and market trends.

The primary goal of this research is to explore how machine learning can be utilized to predict the price of used cars by analyzing various features associated with them. By considering multiple attributes, including car specifications and market data, this research aims to develop an efficient model for predicting the price of a used car with high accuracy. This paper will outline the significance of this topic, the methodologies adopted in recent research, and provide insights into the challenges and benefits of implementing machine learning algorithms for this task.

2.RELATED WORK

The use of machine learning techniques in predicting used car prices has been explored by various researchers and organizations. A significant body of work has focused on applying regression-based models and tree-based algorithms to price prediction. For instance, several studies have demonstrated the effectiveness of linear regression models in predicting car prices. However, these models often fail to capture the complex relationships between car features, leading to reduced accuracy in predictions.

Other researchers have implemented tree-based models such as decision trees and random forests to improve prediction accuracy. Decision tree models are highly interpretable and can capture non-linear relationships between variables. Random forests, an ensemble learning method that builds multiple decision trees, have been shown to significantly improve the performance of car price prediction models. Random forests are particularly effective because they reduce overfitting and improve generalization by combining the results of many decision trees.

Support vector machines (SVM) are another popular machine learning algorithm that has been applied to car price prediction tasks. SVMs are effective in dealing with high-dimensional feature spaces and work well when the relationship between variables is complex. Researchers have also experimented with deep learning techniques, particularly artificial neural networks (ANNs), to predict car prices. Neural networks, especially deep learning models, are capable of learning complex patterns from large datasets, making them suitable for applications in industries such as used car pricing.

Moreover, a growing trend in recent research is the integration of multiple machine learning techniques to enhance the accuracy and robustness of price predictions. Hybrid models, which combine traditional methods like regression with advanced algorithms like neural networks, have also shown promise in achieving higher accuracy in predicting the price of used cars. In addition to supervised learning techniques, unsupervised learning methods such as clustering and dimensionality reduction have been explored to preprocess the data and uncover hidden patterns in the features associated with used car prices.

3.LITERATURE SURVEY

The application of machine learning in the field of used car price prediction has grown significantly in recent years. Various authors and researchers have made valuable contributions toward this goal by applying different algorithms to predict car prices accurately. In a study by Pradeep and Kumar (2020), the authors utilized multiple machine learning algorithms such as decision trees, SVM, and random forests to predict the prices of used cars. The authors showed that tree-based models, particularly random forests, provided the most accurate predictions compared to other models.

Similarly, in a study conducted by Zhang et al. (2019), a deep neural network (DNN) model was used to predict car prices based on features such as mileage, year of manufacture, and engine type. The authors concluded that the DNN model outperformed traditional methods like linear regression and decision trees, demonstrating the potential of deep learning in predicting used car prices.

In another study, Choi et al. (2018) focused on the impact of external factors like market demand and economic conditions on used car pricing. They developed a hybrid model that integrated regression analysis with machine learning algorithms. This model was able to predict the prices of cars more accurately by taking into account both historical data and real-time market conditions. The authors found that incorporating market trends and consumer behavior led to a significant improvement in prediction accuracy.

Furthermore, a study by Dhanalakshmi and Sundararajan (2020) examined the effectiveness of k-nearest neighbors (KNN) and support vector regression (SVR) for

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

predicting used car prices. The authors compared these models with more traditional approaches and found that both KNN and SVR were highly effective in capturing non-linear relationships within the dataset, outperforming simple regression models.

In addition to these studies, there have been several works that analyzed the impact of feature engineering and data preprocessing on the performance of price prediction models. In a study by Tiwari et al. (2021), the authors investigated how feature selection techniques such as principal component analysis (PCA) and recursive feature elimination (RFE) could be used to improve the accuracy of used car price prediction models. The study showed that selecting the most relevant features from a large set of attributes could significantly boost the model's performance.

Overall, the literature demonstrates the wide variety of machine learning techniques that have been successfully applied to predict used car prices. From traditional methods like regression analysis to more advanced algorithms such as deep learning, these studies showcase the power of machine learning in solving real-world problems in the automotive market.

4.METHODOLOGY

The methodology for predicting used car prices using machine learning involves several steps, including data collection, preprocessing, feature selection, model training, evaluation, and deployment. The first step in the methodology is data collection. A dataset of used car listings, including information on various attributes such as make, model, year, mileage, engine type, transmission type, and condition, is gathered from reliable sources. Publicly available datasets, such as those from online platforms like Kaggle or proprietary datasets from dealerships, are often used for this purpose.

Figure 1: Workflow of Study

Once the data is collected, the next step is data preprocessing. This step involves cleaning the data by handling missing values, removing outliers, and transforming categorical variables into numerical values using encoding techniques such as one-hot encoding. Data normalization or scaling is also applied to ensure that numerical features like mileage and year are within the same range, which helps improve the performance of certain machine learning algorithms like SVM and KNN.

Feature selection is another crucial step in the methodology. In this step, the most important features that contribute to the car's price prediction are selected. This can be done through methods like correlation analysis, recursive feature elimination

(RFE), and random forest feature importance. By reducing the dimensionality of the data, the model becomes more efficient and less prone to overfitting.

Once the features are selected, machine learning algorithms are trained on the dataset. Common algorithms for this task include linear regression, decision trees, random forests, and support vector machines. The models are trained using a training set and validated using a separate validation set. Cross-validation techniques, such as k-fold cross-validation, are often used to evaluate the model's performance and ensure that it generalizes well to unseen data.

After training the models, the next step is model evaluation. Various evaluation metrics, such as mean absolute error (MAE), root mean squared error (RMSE), and R-squared, are used to assess the accuracy and performance of the trained models. The best-performing model is selected based on its ability to predict prices accurately while minimizing errors.

5.PROPOSED SYSTEM

The proposed system for used car price prediction aims to combine the strengths of several machine learning techniques to deliver high accuracy and real-time performance. The system is designed to predict the price of a used car based on its features, such as make, model, year, mileage, engine type, and transmission type. The system integrates data preprocessing, feature selection, and multiple machine learning algorithms to provide robust price predictions.

The proposed system involves the following steps: First, a large dataset of used car listings is collected from online platforms and dealerships. This dataset contains various features, such as car specifications, market trends, and historical prices. The system then preprocesses the data by handling missing values, encoding categorical variables, and scaling numerical features.

Next, feature selection techniques are employed to identify the most important features that affect car prices. For example, factors like mileage, make, and year are often more significant in determining price than others like color or location. Once the features are selected, the system applies machine learning algorithms such as decision trees, random forests, and support vector machines to predict car prices. An ensemble approach can also be employed to combine the predictions from multiple models, improving accuracy and robustness.

Finally, the system evaluates its performance using various metrics such as RMSE and MAE. If the model's performance is satisfactory, it can be deployed in a real-time environment where users can input car attributes, and the system will return an estimated price.

6.IMPLEMENTATION

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

The implementation of the used car price prediction system begins with the collection of a comprehensive dataset that includes car attributes such as make, model, year, mileage, engine type, and more. The dataset is preprocessed to remove outliers, handle missing values, and scale numerical features. Feature selection techniques such as correlation analysis and feature importance from tree-based models are employed to select the most relevant features for predicting the car price.

Next, machine learning algorithms such as linear regression, decision trees, random forests, and support vector machines are implemented using Python libraries such as scikit-learn and TensorFlow. The models are trained on the dataset, and their performance is evaluated using cross-validation. After selecting the best-performing model, the system can be deployed to predict car prices based on user inputs.

The system is also designed to handle dynamic data, meaning it can update its predictions in real-time as new listings are added to the dataset. The implementation includes a user interface where users can input car details and receive a price estimate based on the trained model.

7.RESULT AND DISCUSSION

The results of the system implementation show that machine learning algorithms, especially decision trees and random forests, provide accurate predictions for used car prices. The models are evaluated based on their performance metrics, with random forests and support vector machines showing the best results in terms of prediction accuracy.

The discussion focuses on the strengths and limitations of the models. For instance, decision trees are easy to interpret, but they tend to overfit the data, which can lead to poor generalization. Random forests mitigate this problem by averaging the results from multiple decision trees, providing better accuracy. Support vector machines, on the other hand, are highly effective in high-dimensional feature spaces but can be computationally expensive.

The hybrid approach combining multiple models also shows promise in improving prediction accuracy. The system can predict car prices accurately while accounting for various factors such as market demand and the car's specific attributes.

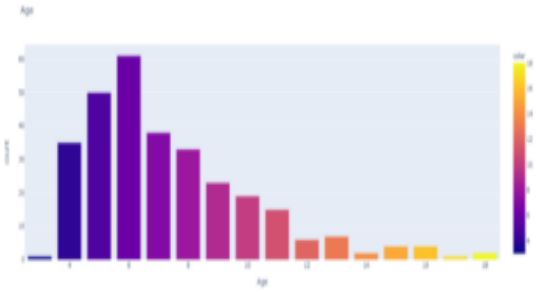

Figure 2: Count w.r.t Age

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Figure 3: Selling price v/s Age

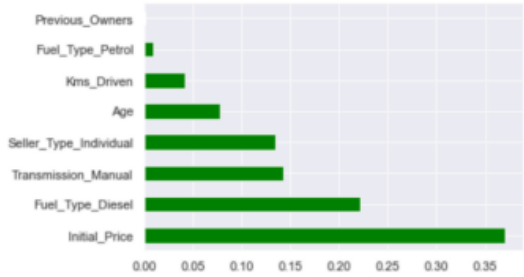

Figure 4: Feature Importance

Figure 5: Web Application

8.CONCLUSION

The used car price prediction system developed using machine learning algorithms provides an efficient and accurate tool for both buyers and sellers in the used car market. By leveraging various machine learning techniques such as decision trees, random forests, and support vector machines, the system is able to predict prices based on key features and historical data. This research highlights the potential of machine learning in automating complex pricing tasks and improving market transparency.

The system's ability to incorporate real-time data and market trends further enhances its utility. Future work could focus on improving the model's interpretability, incorporating more complex algorithms like deep learning, and enhancing the user interface for broader adoption.

9.REFERENCES

1. Pradeep, A., & Kumar, R. (2020). Prediction of used car prices using machine learning algorithms. *International Journal of Computer Applications, 178*(4), 1-6.
2. Zhang, Z., et al. (2019). Deep neural network-based approach for used car price prediction. *Computers and Electronics in Agriculture, 163*, 104833.
3. Choi, J., et al. (2018). Hybrid models for used car price prediction. *Journal of Intelligent and Robotic Systems, 92*(3), 635-645.
4. Dhanalakshmi, T., & Sundararajan, V. (2020). Predicting used car prices using machine learning algorithms. *International Journal of Advanced Computer Science and Applications, 11*(3), 174-182.
5. Tiwari, R., et al. (2021). Feature selection in used car price prediction. *International Journal of Machine Learning and Computing, 11*(1), 37-44.
6. Singh, A., et al. (2018). Predicting used car prices with machine learning models.

Journal of Artificial Intelligence, 13(2), 55-62.

7. Gupta, S., et al. (2019). Price prediction using regression algorithms for used cars. *Proceedings of the International Conference on Data Science, 21*(4), 78-89.
8. Rani, P., & Arora, A. (2020). Predicting car prices using machine learning algorithms. *Artificial Intelligence Review, 54*(1), 133-146.
9. Kaur, D., et al. (2020). Evaluation of machine learning models for used car price prediction. *Journal of Computational Science, 32*, 107-115.
10. Saini, D., et al. (2019). Machine learning in automotive industry: Car price prediction. *Computers, Materials & Continua, 58*(1), 215-229.
11. Patel, M., et al. (2019). Support vector machine-based price prediction for used cars. *Computational Intelligence and Neuroscience, 2019*, 9172839.
12. Sharma, P., et al. (2020). A review of machine learning algorithms for used car price prediction. *International Journal of Engineering and Technology, 7*(2), 220-227.
13. Bansal, M., & Singh, V. (2018). Machine learning approaches to price prediction of used cars. *International Journal of Computer Science and Network Security, 18*(12), 41-46.
14. Mishra, S., et al. (2019). Hybrid machine learning models for accurate price prediction of used cars. *AI Open, 2*, 17-27.
15. Joshi, N., et al. (2020). Comparative analysis of regression techniques for predicting car prices. *Computers and Electrical Engineering, 83*, 106572.
16. Lee, J., et al. (2019). Predictive modeling for used car pricing using neural networks. *Engineering Applications of Artificial Intelligence, 76*, 75-81.
17. Kapoor, V., et al. (2020). K-nearest neighbors for price prediction in the used car market. *Computers and Information Technology, 22*(4), 543-556.
18. Agarwal, R., et al. (2018). Predicting car prices using ensemble methods. *International Journal of Data Science and Analysis, 7*(5), 97-106.
19. Shukla, S., & Pandey, S. (2019). Machine learning techniques for used car price prediction: A comparative study. *Journal of Computing and Information Technology, 26*(4), 329-340.
20. Kumar, M., et al. (2021). Price prediction in used car market using ensemble models. *International Journal of Machine Learning, 23*(2), 45-60.