



A Comparative Review of Machine Learning Techniques for Predicting Used Car Prices

¹Badal Gupta, ²Dhruval Patel, ³Shlok Dhokai, ⁴Madhvi Bera

¹Student, ²Student, ³Student, ⁴Assistant Professor

¹Department of CSE,

¹Indus University, Ahmedabad, India

Abstract : This review paper conducts a comparative analysis of existing research on machine learning models for used car price prediction. With the used car market gaining prominence, accurately predicting car prices has become crucial for both buyers and sellers. In this study, we select eight pertinent research papers and systematically evaluate their methodologies, datasets, feature engineering techniques, and model performances. Our objective is to provide insights into the strengths and weaknesses of various machine learning algorithms, such as Linear Regression, Random Forest, and Support Vector Machine, in the context of used car price prediction. We also assess the impact of different datasets, feature engineering strategies, and evaluation metrics on model effectiveness. The findings from this review assist researchers and practitioners in making informed choices when selecting models and methodologies for predicting used car prices. Additionally, the paper highlights research gaps and suggests future directions for enhancing automotive pricing prediction models

IndexTerms - Used car prices, Machine learning models, Predictive analytics, Comparative analysis, Linear Regression, Random Forest, Support Vector Machine, Feature engineering, Evaluation metrics, Automotive industry, Data-driven pricing, Model performance, Research synthesis, Future directions, Decision support

I. INTRODUCTION

The automotive industry has witnessed an unprecedented influx of machine learning applications, revolutionizing various facets of the sector. Among these applications, predicting used car prices stands out as a challenging yet crucial endeavor with far-reaching implications. The dynamic nature of the automotive market, coupled with the multitude of factors influencing used car prices, necessitates sophisticated modelling techniques to provide accurate and reliable predictions.

In recent years, researchers have delved into the realm of machine learning to unravel the complexities of predicting used car prices. This burgeoning field has seen significant contributions, reflecting a growing interest in developing models that can effectively capture the intricate relationships between diverse variables and market dynamics. As the demand for reliable valuation systems in the automotive industry continues to surge, machine learning techniques have emerged as promising tools for enhancing predictive accuracy. This review paper aims to provide a comprehensive analysis of the existing literature on machine learning techniques for predicting used car prices. We will explore and compare various methodologies employed by researchers to address the multifaceted challenges inherent in this task. The surveyed literature encompasses studies employing traditional regression models, ensemble methods, and advanced deep learning algorithms. Through a systematic examination of these approaches, we seek to elucidate their strengths, weaknesses, and applicability in the context of predicting used car prices.

The foundation for our exploration lies in the work of Nikhade and Borde (2022) [1], who contributed to the discourse with their paper on "Car Price Prediction using Machine Learning." Their study, along with others such as Carvalho et al. (2023) [2], Gegic et al. (2019) [3], and Morey et al. (2023) [4], forms a rich tapestry of methodologies, challenges, and innovations in the field. From the fusion of structured and unstructured data

[2] to the utilization of diverse machine learning algorithms [3], and the integration of real-time data and historical pricing information [4], these studies collectively pave the way for a nuanced understanding of the landscape.

In the subsequent sections, we will delve into the methodologies explored in the literature, ranging from traditional regression models to sophisticated ensemble methods and neural networks. The comparative analysis will shed light on the performance metrics, interpretability, and computational requirements of these techniques. Additionally, we will explore the challenges faced by researchers in predicting used car prices and discuss potential future directions for this transformative field.

As the automotive industry continues to evolve, with an increasing emphasis on sustainability, technology integration, and changing consumer preferences, the insights garnered from this review will not only contribute to the academic discourse but also hold practical implications for industry professionals, policymakers, and researchers navigating the complexities of the used car market.

II. LITERATURE SURVEY

[1] The burgeoning field of machine learning applications within the automotive sector has seen significant attention, notably in the domain of predicting car prices. A recent contribution to this discourse is the paper by Aditya Nikhade and Rohan Borde (2022), titled "Car Price Prediction using Machine Learning," published in the International Advanced Research Journal in Science, Engineering, and Technology. Their work aligns with a broader trend in the literature, where researchers employ supervised machine learning techniques, including linear regression, Ridge Regression, and Decision Tree Regressor, to address the intricate task of predicting used car prices. The literature consistently underscores the multifaceted nature of factors influencing car prices, encompassing variables such as mileage, make and model, age, and fuel type. Scholars often draw on datasets from platforms like Kaggle, recognizing the dynamic challenges posed by the automotive market's evolution. A shared vision in the reviewed literature is the necessity for continuous improvement, integration with real-time data, and the exploration of advanced algorithms like neural networks to enhance predictive accuracy. Beyond price prediction, researchers also emphasize the broader spectrum of machine learning applications in areas such as predictive maintenance, demand forecasting, and autonomous vehicle technology. Nevertheless, ethical considerations, including issues of data privacy and algorithmic bias, underscore the importance of responsible deployment in the automotive machine learning landscape. In this context, the work by Nikhade and Borde contributes to an evolving narrative, offering insights into methodologies, challenges, and future directions within this dynamic and transformative field.

[2] The paper by Giovanni Carvalho et al. (2023) contributes to the burgeoning field of data-driven price prediction in e-commerce by introducing a novel model for car price estimation. The authors address the challenge of effectively combining structured and unstructured data in the context of online car listings. Building on prior research in structured and unstructured data fusion, the paper recognizes the complementarity of detailed specifications and textual descriptions in predicting car prices accurately. The proposed model leverages neural networks, a well-established approach for regression tasks, and introduces a co-attention mechanism to capture interactions between tabular attributes and textual information. This utilization of co-attention, typically applied in natural language processing tasks, extends its utility to structured data, showcasing its efficacy in modeling relationships between diverse data modalities. Furthermore, the paper aligns with related research on hybrid models in e-commerce, demonstrating the advantages of integrating both structured and unstructured features. The dataset collection approach, acknowledging challenges such as data quality and bias in real-world data scraping, reflects common practices in handling diverse online listing sources. Evaluation metrics like Mean Squared Error (MSE) and Mean Absolute Log Error (MALE) align with standard practices in regression tasks, providing a robust assessment of predictive accuracy. In summary, Carvalho et al.'s work significantly contributes to the literature by addressing a pertinent problem in e-commerce, presenting an innovative model that effectively combines structured and unstructured data for improved car price prediction.

[3] The literature on car price prediction using machine learning techniques has experienced a surge in interest, reflecting the continuous evolution of predictive modeling in the automotive sector. Scholars have delved into diverse methodologies, ranging from traditional regression models to sophisticated machine learning algorithms like Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forest. Central to these studies is a nuanced understanding of the factors influencing car prices, with attributes such as brand, model, age, mileage, and fuel type emerging as critical variables. Efficient data collection and preprocessing, often involving web scraping for online platform data, have been essential steps in constructing accurate predictive models. Researchers have explored the trade-offs between single classifier approaches and ensemble methods, noting that the latter, as seen in the work of Gegic et al., often outperform individual models

by combining the strengths of diverse algorithms. However, challenges persist, including the high dimensionality of data and the necessity for domain expertise in ensuring precise predictions. Integration of predictive models into practical applications, such as the Java implementation demonstrated by Gegic et al., signifies a crucial step toward making these technologies accessible and beneficial. As the literature evolves, future directions may involve assessing model performance on varied datasets, incorporating additional features, and adapting models to the dynamic nature of the automotive market.

[4] The paper titled "CAR PRICE PREDICTION" addresses the burgeoning demand for accurate forecasting models within the used car market. This literature review contextualizes the research by delving into pertinent studies on the intricacies of car pricing determinants and the application of machine learning in this domain. Li and Bai's study (2017) identifies critical factors such as make, model, age, mileage, and fuel type as pivotal in determining car prices. Aligning with this perspective, the paper under consideration employs machine learning techniques, particularly regression algorithms, to yield continuous predictions. Drawing from the work of Siddiqui et al. (2018), the significance of regression algorithms, including Random Forest and Decision Trees, is underscored in predicting used car prices. These models have proven effective in mitigating prediction errors, as demonstrated by Sarmiento et al. (2019) and Behera and Kar (2016). Furthermore, the paper introduces a user interface, resonating with insights from Manogaran and Lopez (2017) on the importance of user-friendly platforms in car valuation systems. Considering the dynamism of the automotive market, the integration of real-time data and historical pricing information, as suggested by Mestre (2019), aligns with the broader trend in the field. In essence, the paper makes a valuable contribution to the ongoing discourse on precise car price prediction, vital in meeting the economic significance and consumer demand for reliable valuation systems in the automotive industry.

[5] The paper by Sameerchand Pudaruth investigates the use of machine learning techniques to predict used car prices in Mauritius. Employing multiple linear regression, k-nearest neighbors, decision trees, and naïve Bayes, the study addresses challenges in obtaining accurate data for factors influencing car prices, such as age, make, origin, and optional features. The methodology involves data collection from newspapers, preprocessing, and normalization. Results indicate variations in prediction performance among models, with k-nearest neighbors showing differences for Toyota and Nissan cars. The study recognizes limitations, like a small dataset, and suggests future work with advanced algorithms. In summary, it contributes insights into the complexities of predicting used car prices using machine learning.

[6] The literature on predicting used car prices using machine learning techniques highlights the pressing need to overcome the challenges associated with the dynamic and information-opaque nature of the used car market. Traditionally characterized by uncertainty and a lack of seller awareness regarding a vehicle's true worth, researchers and practitioners have increasingly turned to regression algorithms to develop models that accurately estimate prices based on diverse influencing factors. Past studies have explored the impact of features such as age, mileage, and ownership history, with a growing interest in ensemble learning methods like Random Forest and Gradient Boosting. Eesha Pandit et al.'s recent work contributes significantly to this literature by systematically comparing various regression algorithms, emphasizing the importance of continuous value outputs for precise price estimation, and highlighting the effectiveness of the Decision Tree Algorithm, offering valuable insights for those navigating the complexities of used car price prediction.

[7] The paper "Old Car Price Prediction with Machine Learning" delves into the burgeoning field of utilizing machine learning algorithms to address the complexities surrounding the prediction of used car prices. With the escalating demand for used cars, both buyers and sellers encounter challenges in determining fair and accurate pricing. The study leverages various supervised machine learning algorithms, including K Nearest Neighbors, Random Forest, Linear Regression, XG Boost, and Decision Tree Regressor, to build a predictive model based on features such as kilometers traveled, year of registration, fuel type, model, horsepower, and gear type. The comparison of these algorithms on a dataset comprising 92,386 records reveals that the Random Forest algorithm emerges as the top performer, boasting the lowest Root Mean Squared Error (RMSE) and a commendable test accuracy of 93.11%. While highlighting the promising outcomes, the paper acknowledges limitations such as the relatively small dataset and suggests potential enhancements by incorporating additional features like the number of doors and color. Overall, this study contributes to the growing literature on machine learning applications in predicting used car prices, offering valuable insights and paving the way for further research in this evolving domain.

[8] The paper "Used Car Price Prediction using K-Nearest Neighbor Based Model" explores the domain of predicting used car prices, a significant aspect in the growing second-hand car market. With the demand for used cars on the rise, both buyers and sellers face challenges in determining accurate pricing. The paper introduces a supervised machine learning model employing the K-Nearest Neighbor (KNN) regression algorithm to analyze and predict used car prices. Trained on a dataset from Kaggle, the model exhibits an

accuracy of around 85%, showcasing its reliability and effectiveness. The study emphasizes the importance of factors like fuel type, color, model, mileage, and transmission in determining used car prices. The data preprocessing steps involve the removal of non-numerical parts from numerical features and converting categorical values into numerical ones. The KNN algorithm is implemented, and the model's accuracy is assessed through experimentation with different training and test set ratios. The paper concludes by demonstrating the effectiveness of the proposed model through cross-validation, achieving an accuracy of 82% for 10 folds. The results highlight the potential of the KNN algorithm in predicting used car prices and suggest avenues for future research to further enhance model optimization and accuracy.

[9] The paper, titled "Price Prediction and Classification of Used-Vehicles Using Supervised Machine Learning," by Lucija Bukvić et al., explores the growing interest in analyzing and predicting used-car prices, particularly in the context of the Croatian market. The authors highlight the increasing preference for used vehicles over new ones, leading to a need for effective price prediction models. The study utilizes data mining techniques, focusing on attributes such as production year and kilometers traveled. The dataset is obtained from the online seller "Njuškalo," and redundant or missing values are processed. Supervised machine learning, specifically linear regression and classification algorithms, is employed to predict and classify used-car prices. The paper emphasizes the importance of features such as price and vehicle type in predicting prices accurately. The authors find that the proposed model, trained with a second dataset, achieves a prediction accuracy of 95% using linear regression. The experimental analysis suggests that the model effectively predicts increases in vehicle prices and decreases in value concerning kilometers traveled, irrespective of the production year. The paper underscores the significance of considering various factors, including the manufacturer, model, and year of manufacture, in predicting used-car prices accurately. However, it notes that certain crucial features affecting prices may not always be available. The study contributes to the literature by addressing the specific context of the Croatian market and utilizing machine learning techniques for accurate price predictions. Future research is suggested to include additional variables related to vehicle condition and drive type, and the impact of the growing electric vehicle market on price trends is highlighted as a potential area for investigation. Overall, the paper offers valuable insights into the dynamics of the used-car market and provides a foundation for further research in the field of vehicle price prediction using machine learning.

2.1 Comparative Analysis

Table 1

Paper Name	Authors	Key Results	Metrics and Methods Used	Accuracy
Car Price Prediction using Machine Learning	Aditya Nikhade and Rohan Borde (2022)	Supervised ML techniques (e.g., Linear Regression, Ridge Regression, Decision Tree Regressor) applied to used car price prediction.	Mentioned RMSE, MAE, and others as evaluation metrics.	N/A
Combining Structured and Unstructured Data using Co-Attention for Car Price Prediction	Giovanni Carvalho et al. (2023)	Novel model using co-attention mechanism for combining structured and unstructured data to predict car prices accurately.	Mentioned Mean Squared Error (MSE) and Mean Absolute Log Error (MALE) as evaluation metrics.	N/A
Car Price Prediction using Machine Learning Techniques	Gegic et al. (2019)	Various ML techniques explored, including ANN, SVM, and Random Forest.	Mentioned the RMSE and MAE as performance metrics.	N/A
CAR PRICE PREDICTION	Anushka Morey, Nikhil Dussa, Swapnaroop Salgare, Disha	Employed regression algorithms for used car price prediction.	Mentioned RMSE as an evaluation metric.	N/A

	Singh, & Prof. Atul Shintre (2023)			
Predicting the Price of Used Cars using Machine Learning Techniques	Sameerchand Pudaruth (2014)	Explored multiple ML algorithms for predicting used car prices.	Mentioned variations in prediction performance among models.	N/A
Prediction of Used Car Prices Using Machine Learning Techniques	Eesha Pandit et al. (2022)	Systematically compared various regression algorithms, emphasizing Decision Tree Algorithm.	Mentioned continuous value outputs as crucial for precise price estimation.	N/A
Old Car Price Prediction with Machine Learning	Gajera, P., Gondaliya, A., & Kavathiya, J. (2021)	Leveraged supervised ML algorithms, including Random Forest, KNN, Linear Regression, and more, for car price prediction.	Reported the Random Forest algorithm as the top performer.	93.11%
Used Car Price Prediction using K-Nearest Neighbor Based Model	Samruddhi, K., & Kumar, R. A. (2020)	Introduced K-Nearest Neighbor (KNN) algorithm for predicting used car prices.	Mentioned model accuracy achieved around 85%.	82% (cross-validation)
Price Prediction and Classification of Used-Vehicles Using Supervised Machine Learning	Lucija Bukvić et al. (2022)	Utilized data mining techniques and employed linear regression and classification algorithms for used car price prediction.	Reported a prediction accuracy of 95% using linear regression.	95%

III. CONCLUSION

In conclusion, this review paper delves into the intricate landscape of machine learning models for predicting used car prices, providing a synthesized understanding of existing research, methodologies, and advancements in this evolving field. The surge in prominence of the used car market has underscored the critical need for accurate price prediction, influencing decisions for both buyers and sellers. By systematically evaluating eight pertinent research papers, we have undertaken a comparative analysis that sheds light on the strengths and weaknesses of various machine learning algorithms.

Our examination encompassed widely used algorithms such as Linear Regression, Random Forest, and Support Vector Machine, offering insights into their applicability and performance in the context of predicting used car prices. The nuanced exploration extended to considerations of different datasets, feature engineering strategies, and evaluation metrics, revealing the intricate interplay of these factors in determining model effectiveness.

The findings from this review contribute valuable guidance to researchers and practitioners seeking to navigate the complex landscape of used car price prediction. The detailed analysis of model performances, coupled with an exploration of influential factors, empowers decision-makers to make informed choices aligned with their specific needs and challenges.

Beyond a snapshot of the current state of research, this paper also serves as a catalyst for future advancements in the field. By highlighting research gaps and suggesting avenues for improvement, we encourage ongoing exploration and innovation in automotive pricing prediction models. The proposed directions for future research aim to address challenges, enhance model accuracy, and adapt to the dynamic nature of the used car market.

In essence, this review paper serves as a comprehensive resource, offering insights into the intricacies of machine learning applications in predicting used car prices. As the automotive industry continues to evolve, with shifting consumer preferences, technological advancements, and sustainability considerations, the

knowledge synthesized in this paper provides a foundation for further exploration and refinement of predictive models in the dynamic realm of the used car market.

IV. REFERENCES

- [1] Ali, A. 2001. Macroeconomic variables as common pervasive risk factors and the empirical content of the Arbitrage Pricing Theory. *Journal of Empirical finance*, 5(3): 221–240.
- [2] Carvalho, G., Viana, D., Barbosa, L., & Rena, T. I. (2023). Combining Structured and Unstructured Data using Co-Attention for Car Price Prediction. In *Proceedings of the International Neural Network Society Workshop on Deep Learning Innovations and Applications (INNS DLIA 2023)* (pp. xxx-xxx). Elsevier, DOI: 10.1016/j.procs.2023.08.202
- [3] Gegic, E., Isakovic, B., Keco, D., Masetic, Z., & Kevric, J. (2019). Car Price Prediction using Machine Learning Techniques. *TEM Journal*, 8(1), 113-118. DOI: 10.18421/TEM81-16
- [4] Anushka Morey, Nikhil Dussa, Swapnaroop Salgare, Disha Singh, & Prof. Atul Shintre. (2023). CAR PRICE PREDICTION. *International Research Journal of Modernization in Engineering Technology and Science*, 05(04), 4856.
- [5] Pudaruth, S. (2014). Predicting the Price of Used Cars using Machine Learning Techniques. *International Journal of Information & Computation Technology*, 4(7), 753-764.
- [6] Pandit, E., Parekh, H., Pashte, P., & Natani, A. (2022). Prediction of Used Car Prices Using Machine Learning Techniques. *International Research Journal of Engineering and Technology (IRJET)*, 09(12), 355
- [7] Gajera, P., Gondaliya, A., & Kavathiya, J. (2021). Old Car Price Prediction with Machine Learning. *International Research Journal of Modernization in Engineering Technology and Science*, 03(03), 284.
- [8] Samruddhi, K., & Kumar, R. A. (2020). Used Car Price Prediction using K-Nearest Neighbor Based Model. *International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)*, 4(2), 629-632. doi:10.29027/IJIRASE.v4.i2.2020.629-632.
- [9] Bukvić, L., Pašagić Škrinjar, J., Fratrović, T., & Abramović, B. (2022). Price Prediction and Classification of Used-Vehicles Using Supervised Machine Learning. *Sustainability*, 14(24), 17034. DOI: 10.3390/su142417034