iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

USED CAR PRICE PREDICTION SYSTEM USING MACHINE LEARNING REGRESSION TECHNIQUES

R. Yuvaraj
School of Computing Sciences, VISTAS, Chennai, India
shaheenthoufeeqa@gmail.com

Dr. K. Dharmarajan
Professor, School of Computing Sciences, VISTAS, Chennai, India
dharmak07@gmail.com

ABSTRACT
A Used Car Price Prediction System using machine learning regression techniques aims to provide an accurate, data-driven approach for estimating the resale value of vehicles. The system leverages historical data such as car brand, model, manufacturing year, fuel type, transmission, mileage, engine capacity, and ownership history to train predictive models. Various regression algorithms, including Linear Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting, are implemented and compared to identify the most effective model in terms of accuracy and robustness. Data preprocessing techniques such as handling missing values, encoding categorical variables, and feature scaling are applied to enhance model performance. The dataset is split into training and testing sets to evaluate the predictive capability using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared score. The system also incorporates feature importance analysis to identify key factors influencing car prices. By integrating the trained model into a user-friendly interface, users can input vehicle details and obtain real-time price predictions. This system benefits buyers and sellers by reducing uncertainty and improving transparency in the used car market. Overall, the proposed solution demonstrates how machine learning can be effectively applied to solve real-world pricing problems with improved accuracy and efficiency.

Keywords:
Used Car Price Prediction, Machine Learning, Regression Techniques, Random Forest, Data Preprocessing, Predictive Modelling, Feature Engineering, Price Estimation.

INTRODUCTION

The rapid growth of the automotive industry and the increasing demand for pre-owned vehicles have made accurate price estimation a critical task for both buyers and sellers. A used car price prediction system using machine learning regression techniques aims to address this challenge by providing a data-driven approach to estimate the fair market value of a vehicle. Traditional methods of pricing rely heavily on human judgment, which can often be subjective, inconsistent, and influenced by market fluctuations. In contrast, machine learning offers a systematic and scalable solution by analysing historical data and identifying patterns that influence car prices, such as brand, model, age, mileage, fuel type, transmission, and overall condition. Regression techniques, a fundamental category of supervised learning, play a vital role in predicting continuous numerical values, making them well-suited for price estimation tasks. Algorithms such as linear regression, decision tree regression, random forest regression, and gradient boosting models can be employed to build predictive models with high accuracy. These models learn from large datasets of previously sold cars and establish relationships between various features and their corresponding prices. As a result, the system can generate reliable price predictions for new or unseen data inputs. One of the key advantages of using machine learning in this domain is its ability to handle complex, non-linear relationships among variables, which are often difficult to capture through traditional statistical methods. Additionally, the system can be continuously improved by incorporating new data, thereby adapting to changing market trends and consumer preferences. Feature engineering, data pre-processing, and model evaluation are crucial steps in ensuring the effectiveness and robustness of the prediction system. The implementation of such a system not only enhances transparency in the used car market but also empowers users to make informed decisions. Buyers can avoid overpaying, while sellers can set competitive and realistic prices for their vehicles. Furthermore, this technology can be integrated into online marketplaces, dealership platforms,

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

and mobile applications to provide real-time price recommendations. Overall, a used car price prediction system using machine learning regression techniques represents a practical and innovative solution to modern automotive pricing challenges.

LITERATURE REVIEW

S.No	Author(s) & Year	Title of Paper	Techniques Used	Dataset / Features	Key Findings
1	Bergmann et al., 2025	Machine learning for predicting used car resale prices	Random Forest, Gradient Boosting	Vehicle equipment, mileage, condition	Feature engineering significantly improves prediction accuracy; detailed vehicle attributes enhance model performance
2	Lin et al., 2025	Predicting Used Car Price Based on Machine Learning	Linear Regression, Ridge Regression, Random Forest	CarDekho dataset, mileage, fuel type	Random Forest outperformed linear models in prediction accuracy and robustness
3	Marnholkar et al., 2025	Pre-owned Car Price Prediction using ML	Regression models, data preprocessing	Brand, mileage, transmission, fuel type	Data preprocessing and feature selection play a critical role in improving prediction accuracy
4	Fayyaz et al., 2025	Advanced Feature Engineering in Car Price Prediction	Multiple ML regression models	Accident history, condition, transmission	Emphasizes importance of feature engineering and comparative model evaluation
5	IJCSE (2026)	Car Price Prediction Model using AI	Backward elimination, regression models	Kaggle dataset, correlated variables	Feature selection methods improve model efficiency and reduce overfitting

PROBLEM STATEMENT

A Used Car Price Prediction System using machine learning regression techniques aims to address the challenges faced by buyers and sellers in determining a fair and accurate price for pre-owned vehicles. The used car market is highly dynamic, with prices influenced by multiple factors such as brand, model, manufacturing year, mileage, fuel type, transmission, condition, and market demand. Traditional pricing methods often rely on manual estimation, personal judgment, or inconsistent market listings, which can lead to inaccurate valuations, lack of transparency, and potential financial loss for both parties. Therefore, there is a need for a reliable, data-driven system that can analyse historical data and provide precise price predictions. This project focuses on developing a predictive model using regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, or Gradient Boosting techniques. The system will collect and pre-process large datasets containing various car attributes, handle missing values, and perform feature selection to identify the most influential factors affecting price. By training the model on this data, it can learn patterns and relationships between input features and car prices. The final system will allow users to input specific car details and receive an estimated market price instantly. This not only improves decision-making but also enhances trust and efficiency in the used car marketplace. Additionally, the model can be evaluated using performance metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to ensure accuracy and reliability. Overall, the proposed system leverages machine learning to bring automation, consistency, and fairness to used car price estimation.

OBJECTIVES

The primary objective of the “Used Car Price Prediction System using Machine Learning Regression Techniques” is to develop an accurate, reliable, and efficient model that can estimate the resale value of used cars based on various influencing factors. This system aims to assist both buyers and sellers in making informed decisions by reducing uncertainty and subjectivity in pricing. A key objective is to collect and pre-process a comprehensive dataset containing relevant features such as car brand, model, manufacturing year, fuel type, transmission type, mileage, ownership history, and location. Another objective is to apply and compare multiple regression techniques, such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Support Vector

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Regression, in order to identify the most effective model for predicting car prices. The system also focuses on improving prediction accuracy through feature selection, data cleaning, normalization, and hyper parameter tuning. Additionally, the project aims to evaluate model performance using appropriate metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared values. A further objective is to design a user-friendly interface where users can input car details and instantly receive a predicted price. The system should also be scalable and adaptable to new data, allowing continuous improvement over time. Moreover, the project seeks to demonstrate the practical application of machine learning in solving real-world problems in the automotive market. By achieving these objectives, the system intends to bridge the gap between data-driven insights and everyday decision-making, ultimately enhancing transparency, efficiency, and trust in the used car marketplace.

METHODOLOGY

The methodology for a used car price prediction system using machine learning regression techniques begins with comprehensive data collection from reliable sources such as online car marketplaces, dealerships, and automotive databases. The dataset typically includes features like car brand, model, manufacturing year, fuel type, transmission type, mileage, engine capacity, ownership history, and location. Once the data is gathered, the next step involves data pre-processing, where missing values are handled through imputation or removal, duplicate entries are eliminated, and categorical variables are encoded using techniques such as one-hot encoding or label encoding. Numerical features are scaled or normalized to ensure uniformity and improve model performance. Exploratory data analysis (EDA) is then performed to identify patterns, correlations, and outliers using statistical summaries and visualization tools. Following pre-processing, feature selection and engineering are applied to improve the predictive power of the model. Irrelevant or redundant features are removed, and new meaningful features may be derived, such as car age from the manufacturing year or price per kilometre. The dataset is then split into training and testing sets, typically in an 80:20 ratio, to evaluate model performance effectively. Various regression algorithms are implemented, including Linear Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting methods. Each model is trained on the training dataset, and hyper parameter tuning is performed using techniques such as grid search or cross-validation to optimize performance. Model evaluation is conducted using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) score to compare the accuracy of different models. The best-performing model is selected based on these metrics. Once the optimal model is identified, it is validated using the test dataset to ensure its generalization capability. Finally, the trained model is deployed in a user-friendly interface, such as a web application, where users can input car details to receive an estimated price. Continuous monitoring and periodic retraining of the model are recommended to maintain accuracy as market trends and pricing dynamics evolve over time.

SYSTEM ARCHITECTURE

The system architecture for a Used Car Price Prediction System using machine learning regression techniques is designed as a multi-layered framework that ensures efficient data processing, model training, and user interaction. The architecture begins with the data acquisition layer, where historical used car data is collected from various sources such as online marketplaces, dealership records, and datasets. This data typically includes features like car brand, model, manufacturing year, mileage, fuel type, transmission type, location, and selling price. The collected raw data is then passed to the data pre-processing layer, where it undergoes cleaning, handling of missing values, encoding of categorical variables, normalization, and feature selection to improve data quality and model performance. Next, the processed data is fed into the machine learning layer, where regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, or Gradient Boosting are applied. This layer involves splitting the dataset into training and testing sets, model training, hyper parameter tuning, and evaluation using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared score. The best-performing model is selected and stored for deployment. The deployment layer integrates the trained model into a web or mobile application through APIs, enabling real-time predictions. The user interface layer allows users to input car details such as age, mileage, and specifications. These inputs are processed and sent to the model, which returns an estimated price instantly. Finally, the system includes a feedback and monitoring component to track prediction accuracy and update the model periodically with new data, ensuring continuous improvement and reliability of the system.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

IMPLEMENTATION DETAILS

The system architecture for a Used Car Price Prediction System using Machine Learning Regression Techniques is designed as a multi-layered framework that ensures efficient data processing, model training, and user interaction. At the core, the architecture begins with the data collection layer, where historical data about used cars is gathered from multiple sources such as online marketplaces, dealership databases, and user inputs. This data typically includes features like car brand, model, manufacturing year, mileage, fuel type, transmission, and location. The collected raw data is then passed to the data pre-processing layer, where it undergoes cleaning, transformation, and normalization. Missing values are handled, categorical variables are encoded, and irrelevant or noisy data is removed to ensure high-quality input for the model. Next, the processed data is fed into the **f**eature engineering and selection layer, where important features are identified and optimized to improve prediction accuracy. Techniques such as correlation analysis and dimensionality reduction may be applied to enhance model performance. This refined dataset is then used in the model training layer, which is the core component of the system. Various regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting are trained and evaluated. The system selects the best-performing model based on evaluation metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared value. Once the optimal model is finalized, it is deployed in the model serving layer, where it becomes accessible for real-time predictions. This layer is typically integrated with an API that allows communication between the backend model and the frontend interface. The user interface layer enables users to input car details through a web or mobile application. When the user submits the details, the system processes the input, sends it to the trained model, and returns the predicted price instantly. Finally, the architecture includes a monitoring and feedback layer, which continuously tracks model performance and user interactions. This layer allows periodic retraining of the model using new data, ensuring that the system remains accurate and up-to-date with market trends. Overall, this architecture ensures scalability, accuracy, and efficiency in predicting used car prices.

RESULT & PERFORMANCE EVALUATION

The system architecture for a Used Car Price Prediction System using machine learning regression techniques is designed as a multi-layered framework that ensures efficient data flow, processing, and prediction. The architecture begins with the data collection layer, where large volumes of historical used car data are gathered from various sources such as online marketplaces, dealership databases, and public datasets. This data typically includes attributes like car brand, model, manufacturing year, mileage, fuel type, transmission, ownership history, and location. The collected raw data is then passed to the data pre-processing layer, where it undergoes cleaning, handling of missing values, removal of outliers, and transformation into a suitable format. Categorical variables are encoded, and numerical features are normalized or standardized to improve model performance. Following pre-processing, the refined dataset is stored in a structured database or data warehouse, forming the data storage layer. The next stage is the feature engineering and selection layer, where relevant features are extracted and optimized to enhance prediction accuracy. Techniques such as correlation analysis and dimensionality reduction may be applied to identify the most influential variables. The processed data is then fed into the machine learning model layer, where various regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting are trained and evaluated. Model training involves splitting the dataset into training and testing sets, followed by performance evaluation using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared. Once the optimal model is selected, it is deployed in the prediction layer, which serves as the core of the system. This layer accepts user inputs through a user interface, such as a web or mobile application, where users enter car details. The input data is processed in real time and passed to the trained model to generate price predictions. The system also includes an API layer to facilitate communication between the frontend and backend components. Finally, the output layer displays the predicted price to the user in an intuitive format. The architecture may also include a feedback and retraining mechanism to continuously improve model accuracy by incorporating new data over time.

ADVANTAGES & LIMITATIONS

The system architecture for a **Used Car Price Prediction System using Machine Learning Regression Techniques** is designed as a multi-layered pipeline that ensures efficient data processing, model training, and user interaction. The architecture begins with the **data collection layer**, where historical used car data is gathered from various sources such as online car marketplaces, dealership databases, and public datasets. This data typically includes attributes like car brand, model, manufacturing year, fuel type, mileage, transmission type, and price. The collected data is then passed to the **data pre-processing layer**, where it undergoes cleaning, handling of

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

missing values, outlier removal, and transformation. Categorical variables are encoded, and numerical features are normalized or scaled to improve model performance. Next, the processed data moves to the **feature engineering and selection layer**, where relevant features are extracted and optimized to enhance predictive accuracy. The refined dataset is then fed into the **machine learning layer**, where various regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting are trained and evaluated. Model selection is performed based on performance metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared value. Once the optimal model is selected, it is deployed in the **model serving layer**, which exposes prediction functionality via APIs. The **application layer** provides a user-friendly interface where users can input car details and receive estimated prices in real time. Finally, the **monitoring and feedback layer** tracks system performance and collects user feedback to continuously improve the model through periodic retraining, ensuring accuracy and adaptability to market trends.

FUTURE SCOPE

The future scope of a used car price prediction system using machine learning regression techniques is vast and promising, driven by advancements in data availability, computational power, and intelligent algorithms. In the coming years, such systems can be enhanced by integrating real-time data from multiple sources, including market trends, economic indicators, fuel prices, and consumer demand patterns, leading to more accurate and dynamic pricing models. The incorporation of advanced techniques like deep learning and ensemble methods can further improve prediction accuracy by capturing complex nonlinear relationships among variables. Additionally, the system can be expanded to include image processing capabilities, allowing users to upload car images for automated condition assessment using computer vision. Integration with IoT-enabled vehicle data, such as mileage tracking and maintenance history, can also refine predictions. Another potential direction is the development of personalized pricing models that consider user preferences, location, and buying behavior. Deployment as a cloud-based or mobile application can enhance accessibility and scalability for both buyers and sellers. Furthermore, incorporating explainable AI techniques will make predictions more transparent and trustworthy. In the long term, such systems could be integrated into online marketplaces and dealership platforms, enabling automated negotiations and smarter decision-making, ultimately transforming the used car trading ecosystem into a more efficient and data-driven marketplace.

CONCLUSION

The used car price prediction system developed using machine learning regression techniques demonstrates the practical application of data-driven approaches in solving real-world problems. By leveraging algorithms such as Linear Regression, Decision Trees, and Random Forest, the system effectively analyses multiple factors including vehicle age, mileage, fuel type, transmission, and brand to estimate accurate market prices. The project highlights the importance of data pre-processing, feature selection, and model evaluation in achieving reliable predictions. Among the models tested, ensemble methods like Random Forest often provided better accuracy due to their ability to handle non-linear relationships and reduce overfitting. The system not only improves transparency in pricing but also assists buyers and sellers in making informed decisions, thereby reducing the chances of fraud and price manipulation in the used car market. Furthermore, the integration of such predictive systems into online platforms can significantly enhance user experience by providing instant and data-backed price recommendations. Despite its effectiveness, the model's performance depends heavily on the quality and diversity of the dataset, indicating the need for continuous data updates and model retraining. Future improvements may include incorporating advanced techniques such as deep learning, real-time data integration, and region-specific pricing adjustments. Overall, this project successfully demonstrates how machine learning regression techniques can be utilized to build intelligent, scalable, and practical solutions for the automotive resale industry.

REFERENCES

1) S. Vandhana, S. Nithish Kumar, V. Namo and A. Ragul Prasad, "Car Price Prediction and Recommendation System Using Machine Learning," in Proc. Int. Conf. Innovations in Computational Intelligence and Computer Vision (ICICV), 2025, pp. 533–546, 2026.
2) R. Chen, "Car Price Prediction Using Machine Learning," in Proc. Int. Conf. E-commerce and Artificial Intelligence (ECAI), 2024, pp. 536–541, 2025.
3) V. Venugopal and S. Haseena, "Accurate Price Prediction of Used Cars Using Machine Learning Models," 2025.
4) Y. Zheng, "Machine Learning Optimization and Challenges in Used Car Price Prediction," ITM Web Conf., 2025.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

5) T. Li, "Used Car Price Prediction and Feature Importance Analysis Using XGBoost and SHAP," Highlights in Business Economics and Management, 2025.
6) C. Li, "Machine Learning-Based Models for Accurate Car Prices Prediction," Expert Systems Applications, 2025.
7) S. Bergmann and S. Feuerriegel, "Machine Learning for Predicting Used Car Resale Prices Using Granular Vehicle Equipment Information," Expert Systems with Applications, 2025.
8) M. Faraz Ahmad et al., "Car Price Prediction Using Machine Learning," 2025.
9) A. Kandula et al., "Refurbished Car Price Prediction Using Machine Learning," in Proc. Int. Conf. Sustainable Communication Networks and Applications, 2025.
10) R. Nuzulia et al., "Comparative Analysis of XGBoost and Random Forest for Used Car Price Prediction," in Proc. Int. Conf. Electrical Engineering and Informatics, 2025.
11) J. S. Jhala and D. Anand, "Comparative Analysis of Supervised Learning Algorithms for Valuating Used Car Prices," in Proc. Int. Conf. Advancement in Computation & Computer Technologies, 2025.
12) A. R. Lakamsani et al., "Machine Learning-Based Used Car Price Prediction System," IEEE Access, 2025.
13) H. Zhang, "Prediction of Used Car Price Based on LightGBM," in Proc. Int. Conf. Advanced Electronic Materials, Computers and Software Engineering, 2025.
14) Y. Li, Y. Li and Y. Liu, "Research on Used Car Price Prediction Based on Random Forest and LightGBM," in Proc. IEEE Int. Conf. Data Science and Computer Application, 2025.