Grenze International Journal of Engineering and Technology, January Issue

Car Price Prediction using Machine Learning Techniques

Lokesh Khedekar[1], Sayyam Jain[2], Sakshi Jadhav[3], Saurabh Jagdale[4], Jaina Jain[5] and Janvi Jain[6]
[1-6]Viswakarma Institute of Technology, Pune, Maharastra, India
Email: lokesh.khedekar@vit.edu, sayyam.jain24, sakshi.jadhav24, saurabh.jagdale24, jaina.jain24, janvi.jain24@vit.edu

***Abstract*— Accurate estimation of used car prices is pivotal in the automotive industry, influencing decisions for buyers, sellers, and financial institutions. Traditional valuation methods often rely on subjective assessments, leading to inconsistencies. This study introduces a data-driven approach employing machine learning techniques to predict used car prices with enhanced precision. Utilizing a comprehensive dataset encompassing various car attributes, we implemented and compared multiple regression models, including Linear Regression, Decision Tree, Random Forest, and XGBoost. The XGBoost Regressor outperformed others, achieving an R^2 score of 0.8630, indicating its robustness in capturing complex patterns within the data. To facilitate user interaction, the model was deployed through a Flask-based web application, featuring an intuitive interface for real-time price predictions. The integration of machine learning with web technologies offers a reliable tool for stakeholders, streamlining the car valuation process and promoting informed decision-making in the used car market.**

***Keywords*— Car Price Prediction, XGBoost, Machine Learning, Web Application, Flask, Vehicle Analytics, Regression Model, Random Forest.**

I. Introduction

In the automotive industry, data-driven decision-making has witnessed a shift toward the adoption of advanced technologies in solving complex problems. A challenge faced not only by individual car buyers but also by stakeholders in the industry is an accurate estimation of car prices. Factors that influence the valuation of a vehicle include brand, model, mileage, engine type, fuel efficiency, year of manufacture, and demand in the market. Traditional methods for determining price have always relied on subjective assessment or a generalized price book, which allows room for inconsistency and in accuracy.

This set the stage for more robust data-centric approaches to enter, particularly those based on machine learning. Machine learning, a branch of artificial intelligence, helps reveal hidden regularities within large datasets and build predictive models based on those patterns. Using ML methodologies, one can build systems that learn from historical car sales data and henceforth predict future car prices quite accurately. The adaptability of machine learning models to new data and their capacity to generalize across diverse scenarios make them a good fit for dynamic, heterogeneous markets like automobiles. Hence, this work, titled "Car Price Prediction Using Machine Learning Techniques," undertakes the exploration and evaluation of various ML algorithms in predicting the selling price of used cars, with the ultimate goal of minimizing the prediction error as well as making the pricing mechanism more reliable.

Grenze ID: 01.GIJET.12.1.538_1

This work uses a large dataset having information on various attributes of cars and applies multiple algorithms—Linear Regression, Decision Trees, Random Forest, and XGBoost—to compare their performance metrics and pick the best efficient model. The impetus for this research arises from the very real difficulties that consumers experience in ascertaining the reasonable price of a pre-owned vehicle. Most buyers and sellers do not have the requisite technical expertise or acumen to estimate the value of a car; hence, many end up incurring financial losses or remaining dissatisfied with their purchases.

II. LITERATURE REVIEW

- The study deals with multiple machine learning algorithms such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forests (RF). After comparative analysis of these machine learning algorithms, the performance of the Random Forest model is superior in terms of accuracy. It has some limitations, like ANN and SVM have limited interpretability, and hyperparameter tuning is not clearly addressed, which may affect overall model performance.
- The study explores traditional machine learning algorithms including Linear Regression, Multiple Linear Regression, Decision Tree, and Random Forest to predict car price. Despite their strengths, it also has some limitations, like ROC-AUC is mentioned but not used for evaluation, so the evaluation framework is unclear. Some techniques like clustering and logistic regression are mentioned without clear relevance.
- The study used a number of machine learning algorithms in which Linear Regression, Decision Tree Regressor, Random Forest Regressor, and Gradient Boosting Regressor were included for car price prediction, in which Random Forest Regressor got the highest accuracy. The overall system is implemented in Python. The model performance was evaluated using standard regression methods including R^2, MAE, MSE, and RMSE. However, the implemented methodology has several limitations, like the model is structured by taking the Indian market at center, so the study becomes regional-specific, which may restrict generalizability.
- Car price prediction involves various factors to estimate the fair market value of a vehicle. This study involves various machine learning techniques including Linear Regression, Random Forest, XGBoost Regressor, and Neural Network, among which XGBoost gives the highest accuracy. The machine learning models were evaluated based on R^2 and MAPE evaluating methods. But poor data quality leads to inaccurate car price prediction, which is a limitation of this study.
- The research paper explores the applications of various machine learning techniques for prediction of car price by comparative analysis of various algorithms. The limitation of the study is that it didn't summarize findings broadly with detailed metrics like R^2 and MAPE, and also how the missing or noisy data is handled is not discussed clearly.
- The study uses different machine learning techniques—Random Forest, Support Vector Machine, Logistic Regression, and Linear Regression—for comparative study. It applies optimized BP neural network technique for vehicle price prediction. The limitation is that it uses a lot more computer power than single machine learning algorithms.

III. METHODOLOGY

- Data Collection: Our research used a comprehensive dataset from Kaggle, which is in CSV format, for preprocessing the data. The collected dataset is used as the foundation for training and testing various predictive models.
- Data Processing: We implemented various data preprocessing techniques such as imputation methods for handling missing values and data cleaning for removing outliers and inconsistencies to improve the quality of the dataset.
- Feature Selection: The features included in the dataset are brand, model, transmission, owner, and fuel type, which are categorical, whereas age and kmDriven are numerical. Here, age is a new numerical feature derived by subtracting the year of manufacture from the year 2025, and the target variable, 'AskPrice,' represents the seller's listed price for the used car, which our model predicts based on input features.
- Model Selection and evaluation: We trained and compared four regression models: Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, and XGBoost Regressor. The coefficient of determination (R^2 score) method was used to assess the performance of the models. The R^2 score measures the proportion of variance in the target variable explained by the model. The XGBoost Regressor achieved

the highest performance with an R^2 score of 0.8630, i.e., approximately 86.3% variance explained in car price. As a result, the XGBoost Regressor model was selected for car price prediction.
- Model Deployment: The trained XGBoost Regressor model was serialized using joblib and integrated into a Flask-based web application. For user interaction and administrative access, HTML templates were created.

IV. Implementation

Technical integration of the machine learning model with a responsive web interface using a Flask backend and the use of HTML, CSS, and JavaScript for the frontend is outlined in this section. The aim is to provide a user-friendly car price predicting system for both buyers and system administrators.

A. Backend Development

The Flask framework in Python, which was ideal for its simplicity and ease of integration with machine learning workflows, is used to build the backend. The prediction engine is powered by a trained XGBoost Regressor model. For deployment of the model, the trained model is saved using the joblib library so that it can be easily loaded without requiring retraining. When a user submits their car details through the website, the backend takes the following steps:
- It reads and processes the input data from the form.
- The data is then transformed using a previously saved Column Transformer, ensuring it matches the format used during training.
- The cleaned data is fed into the XGBoost model to predict the car's price.
- The system logs the inputs and prediction for future analysis.
- Finally, the predicted price is sent back and displayed to the user.
  - Two main routes handle the backend logic:

/predict — processes user submissions and shows results.
/admin — gives system administrators access to all past predictions in one place.

B. Frontend Design

The web interface is designed using HTML, with three pages (index.html, admin.html, result.html), CSS for styling, and JavaScript for responsiveness.
- Index.html (User Input Page): This is the homepage for the user. Using the Web Speech API, it provides voice input support for the brand and model fields, making it easier for users to interact with the system hands-free. Based on the age of the car, the system gives a quick assessment of the car's condition. Users can view pricing trends through an embedded Chart.js chart and switch between light and dark modes forcomfort.
- Admin.html (Admin Dashboard): This page is designed for administrators to monitor system activity. It shows a table of all past predictions, calculates statistics like average price and most predicted brand, and offers an option to export the data to CSV forfurtheranalysis.
- Result.html (Prediction Page): After submitting user details, the predicted price is shown alongside key information such as the car's make, model, kilometers driven, and condition. A bar chart illustrates which features had the biggest impact on the prediction. The page also displays the current time and supports dark mode for a better viewing experience.

For connecting all the pages, Jinja2 templating is used, which allows dynamic content to be displayed seamlessly. The result is a fully functional, interactive web application that makes AI-powered car price prediction both accessible and informative.

V. Result and Discussion

The four regression models—Linear Regression, Random Forest, Gradient Boosting, and XGBoost—were evaluated for predictive performance using the R^2 score, which is commonly utilized to assess the correctness of regression models.

As can be observed from Figure 1 the Linear Regression model provided an R^2 value of 0.5373, indicating a moderate fit to the data. Performance was significantly improved with ensemble models.

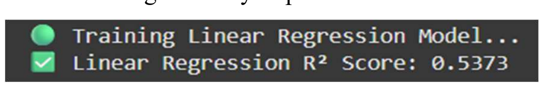


Figure 1: Linear Regression model Accuracy (R^2: 0.5373)

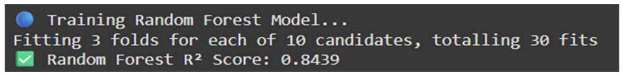


Figure 2: Random Forest model Accuracy (R^2: 0.8439).

Figure 2 shows the Random Forest model providing an R^2 value of 0.8439.

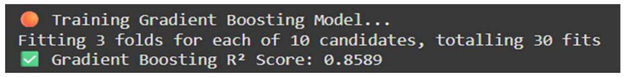


Figure 3: Gradient Boosting Model Accuracy (R^2: 0.8589).

Figure 3 also shows the Gradient Boosting model output with the R^2 value of 0.8589.

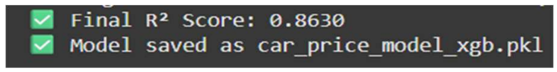


Figure 4: Final R^2 Score: 0.8630. Model saved as car_price_model_xgb.pkl

The best performance was achieved by XGBoost, as shown in Figure 4, with an R^2 value of 0.8630.
To enable an equal comparison, the R^2 scores of all models were presented side by side in a bar graph (Figure 5). The graph clearly demonstrates how much improvement is possible using sophisticated ensemble methods, with the best performance obtained by XGBoost.

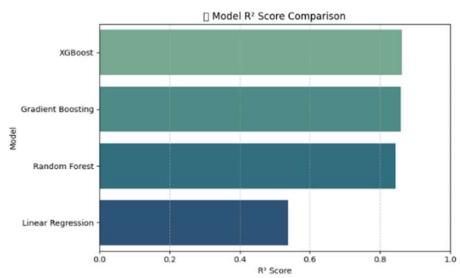


Figure 5: Bar graph comparing R^2 scores of all models.

A. User Interface and Experience

The web interface of the Car Price Prediction System has been designed specifically to offer a seamless user experience. Below are some of its key features:
Interactive Input Fields: Users can enter car information (brand, model, transmission type, etc.), which is then processed on the backend to generate predictions.
Real-time Prediction: Users are presented with an instant predicted price as soon as the information is submitted, along with insights into the main factors influencing the prediction.
Dark Mode Toggle: For ease of use, a dark mode toggle is available to enhance user comfort during prolonged usage.
Figure 6 shows the input form through which car details can be entered to predict car prices.

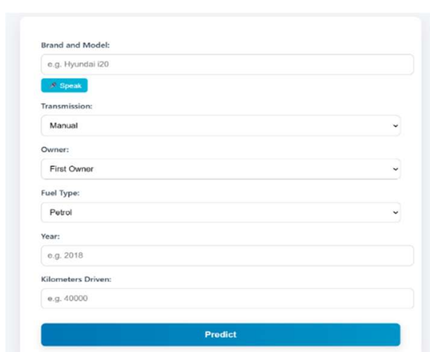


Figure 6 (Car Price Prediction Form): "User interface for inputting car details (brand, model, mileage, etc.)."

The page shows where users input information like brand, model, fuel type, and mileage.

B. Presentation of Prediction Results and Explanation

Once the user provides car details, the estimated price is displayed along with the features that most influenced the prediction. This helps users understand the reasoning behind the estimated price.
Figure 7 is the prediction result page upon which the predicted price and the reasons for prediction are emphasized.

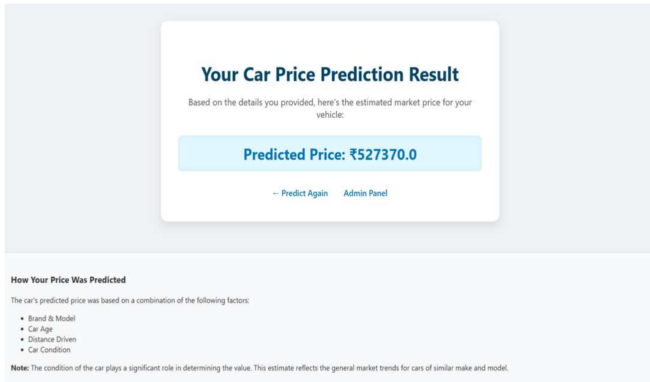


Figure 7 (Prediction Result Page): "Displays predicted price and key influencing factors (e.g., brand, mileage)."

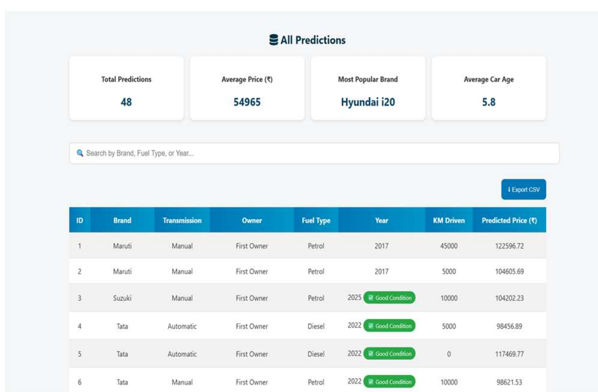
The price prediction is displayed along with the reasons like brand, model, and mileage for the prediction.

C. Admin Panel Health Condition Management and Suggestions

The admin panel provides monitoring and management of data obtained from users. The system keeps records of the following for each prediction:

- ➢ Number of predictions made.
- ➢ Average predicted cost, age, and mileage.
- ➢ Most predicted models and brands.

This information is presented in an exportable, interactive, filterable table that administrators can export and filter as CSV for further analysis. An admin activity snapshot is depicted in Figure 5 (if this refers to the admin dashboard, please ensure figure numbering is unique).



ID	Brand	Transmission	Owner	Fuel Type	Year	KM Driven	Predicted Price (₹)
1	Maruti	Manual	First Owner	Petrol	2017	45000	122596.72
2	Maruti	Manual	First Owner	Petrol	2017	5000	104605.69
3	Suzuki	Manual	First Owner	Petrol	2025 Good Condition	10000	104202.23
4	Tata	Automatic	First Owner	Diesel	2022 Good Condition	5000	98456.89
5	Tata	Automatic	First Owner	Diesel	2022 Good Condition	0	117469.77
6	Tata	Manual	First Owner	Petrol	2022 Good Condition	10000	98621.53

Figure 8 (Admin Dashboard): "Administrative view showing prediction history, average price, and popular brands."

The admin dashboard (Figure 8) provides a tabular view of all predictions, average price statistics, and export functionality for further analysis.

D. FAQ Integration: Enhancing User Trust and Usability

To ensure transparency, the system includes a FAQ module on the result page that addresses common user concerns. This component was specifically designed to improve user understanding, trust, and interaction with the system.

Highlights from the FAQ

- Prediction Methodology Explained: Users are informed that predictions are based on historical and multiple factors including car brand, model, mileage, fuel type, and age.
- Saving Predictions: Users are provided with an option to save their predictions upon login, supporting convenience and profile-based tracking.
- Accuracy Awareness: While the system achieves high accuracy, users are cautioned that predictions may vary depending on the input quality and completeness.
- Regular Data Updates: The prediction engine relies on frequently updated datasets to align with market trends, ensuring estimates reflect real-time value shifts.
- Factors Considered: Transparency is promoted by listing out the factors influencing predictions — such as owner history, fuel type, and transmission — reinforcing the depth of model input.

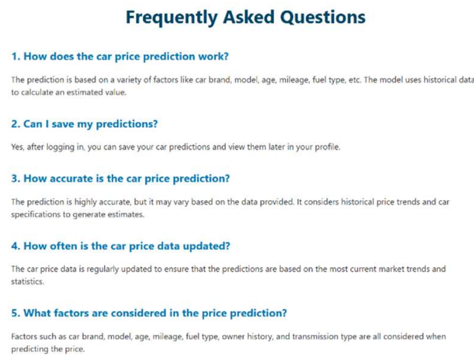


Figure 9: FAQ Page.

VI. CONCLUSION

This study describes a system using machine learning to predict used car prices accurately. It uses organized data and advanced prediction models. While the XGBoost model achieved high accuracy (R^2: 0.8630), future work could explore larger datasets or hybrid models to improve generalizability across global markets. Important steps included preparing the data carefully, creating new features like car age and brand-model pairs, and fine-tuning settings. The final model was put into use with a Flask web app. This app has a simple, easy-to-use design showing predictions and charts instantly. This approach shows how machine learning and web technology can help people make smart choices in the used car market.

REFERENCES

[1] Y. S. Balcioglu and B. Sezen, "Car price prediction using machine learning techniques," in Proceedings of the 6th International Artemis Congress on Health and Sport Science, Mar. 2024.

[2] A. R., "Car price prediction using machine learning techniques," International Research Journal of Modernization in Engineering Technology and Science, vol. 4, pp. 45–50, 2022.

[3] V. K. H., S. V., A. V., and S. Srinivas, "Comparative analysis of machine learning algorithms for used car price prediction," International Journal of Current Science Research and Review, vol. 7, no. 9, pp. 7220–7228, Sep. 2024.

[4] A. Zhu, "Pre-owned car price prediction using machine learning techniques," in Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023), pp. 356–360, 2024.

[5] S. Gupta and S. K. S., "Car price prediction using machine learning," International Journal of Research Publication and Reviews, vol. 5, no. 1, pp. 4758–4762, Jan. 2024.

[6] O. Sampath and A. G. Anusha, "Prediction of used car price using artificial neural networks and machine learning," International Journal of Engineering Science and Advanced Technology, vol. 24, no. 5, pp. 373–378, May 2024.

[7] K. Madhusudhanan et al., 'ProbSAINT: Probabilistic Tabular Regression for Used Car Pricing,' arXiv:2403.03812, 2024.

[8] A. A. Rehman, M. M. Rathore, et al., "A machine learning-based approach for predicting car prices: A case study on used cars," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 10, no. 9, pp. 613–618, 2019.

[9] N. Pal et al., 'Predicting Used Cars Prices Using Random Forest,' arXiv:1711.06970, 2017.

[10] R. R. Yang, S. Chen, and E. Chou, "AI Blue Book: Vehicle price prediction using visual features," arXiv preprint, arXiv:1803.11227, Mar. 2018.