ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

PREDICTING USED CARS PRICES USING MACHINE LEARNING

J.Venkataramana[1], U.Sashi[2], L.Haritha[3], B.Pavani[4], P.Sravani[5], A.Lakshmi Sravanthi[6]

[1]Assistant Professor, Dept of CSE, Gouthami Institute of Technology and Management For Women, Andhra Pradesh, India
[2 3 4 5 6] U.G Student, Dept of CSE, Gouthami Institute of Technology and Management For Women, Andhra Pradesh, India

ABSTRACT

The used car market is complex, with prices influenced by factors like age, mileage, brand, condition, and market trends, making accurate price prediction a challenge for both buyers and sellers. This project proposes a machine learning-based system to predict used car prices by analyzing historical sales data and key features, using algorithms such as Linear Regression, Random Forest, and Gradient Boosting to provide reliable price estimates. The study explores various regression techniques, including Linear, Decision Tree, SVM, Neural Networks, and Bagged Trees, to assess their effectiveness in predicting used car prices. Evaluation metrics will be used to compare models based on their performance and error rates, with deep neural networks showing exceptional results due to low RMSE and MSE values. Some models, like cubic and fine Gaussian SVMs and wide neural networks, demonstrate strong correlations between input and output variables. Ultimately, Bagged Trees are identified as the most cost-effective option, balancing performance and pricing for accurate predictions.

INTRODUCTION

The automotive industry, especially the used car market, has seen rapid growth due to increasing demand for affordable transportation, making accurate and transparent pricing essential. Pricing a used car involves numerous factors such as make, model, mileage, condition, and market trends, which makes traditional methods inadequate. To address this, machine learning (ML) is being leveraged to predict prices with greater accuracy and consistency, using algorithms that can model complex relationships in large datasets. This project uses supervised ML techniques to develop a robust system for predicting used car prices, exploring algorithms like Linear Regression, Random Forests, and XGBoost. By incorporating features like vehicle age, mileage, and market trends, and applying data preprocessing, hyperparameter tuning, and cross-validation, the system aims to provide reliable price estimates. The final product will be a user-friendly application that integrates real-time data, making it useful for buyers, sellers, dealerships, and financial institutions. However, challenges such as data quality, model interpretability, and generalizability across regions must be addressed to ensure the system's effectiveness and adaptability in diverse markets.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Literature Review:

The used car market has seen rapid growth over the past decade, prompting extensive research into accurate price prediction using machine learning (ML). Early studies focused on traditional regression models but were soon surpassed by advanced techniques like Decision Trees, Random Forests, SVMs, Gradient Boosting (e.g., XGBoost, LightGBM), and Neural Networks, which better captured non-linear patterns and complex feature interactions. Key vehicle attributes—such as make, model, mileage, year, fuel type, and transmission—emerged as critical predictors. Research also highlights the importance of feature engineering, data preprocessing, and model interpretability, with tools like SHAP and LIME enhancing transparency. Comparative studies favor ensemble models for their robustness, though model performance varies by dataset. Hybrid approaches that combine ML with domain knowledge show promise, especially in data-scarce regions. Challenges like data imbalance, concept drift, and regional bias remain, driving interest in adaptive, real-time, and user-friendly systems with interactive interfaces and explainable outputs for practical deployment.

EXISTING METHOD

The task of used car price prediction has evolved from traditional heuristic and rule-based methods to more sophisticated data-driven techniques, including linear regression, polynomial regression, decision trees, and ensemble learning methods like Random Forest and Gradient Boosting. While earlier models were limited by their inability to capture complex, non-linear relationships, modern approaches such as Support Vector Regression, Artificial Neural Networks, and hybrid models like stacking have significantly improved accuracy. However, these methods still face challenges, including overfitting, interpretability, and the need for large datasets and computational power. Feature engineering and data enrichment techniques, as well as clustering algorithms, have further enhanced prediction accuracy by uncovering underlying market structures. Despite the advances, limitations such as data imbalance and the dynamic nature of the market persist, requiring ongoing innovation in both models and data processing pipelines.

PROPOSED METHOD

The task of used car price prediction has evolved from traditional heuristic and rule-based methods to more sophisticated data-driven techniques, including linear regression, polynomial regression, decision trees, and ensemble learning methods like Random Forest and Gradient Boosting. While earlier models were limited by their inability to capture complex, non-linear relationships, modern approaches such as Support Vector Regression, Artificial Neural Networks, and hybrid models like stacking have significantly improved accuracy. However, these methods still face challenges, including overfitting, interpretability, and the need for large datasets and computational power. Feature engineering and data enrichment techniques, as well as clustering algorithms, have further enhanced prediction accuracy by uncovering underlying market structures. Despite the advances, limitations such as data imbalance and the dynamic nature of the market persist, requiring ongoing innovation in both models and data processing pipelines.

SOFTWARE REQUIREMENTS:

- **Operating System:** Windows 8 and above
- **Coding Language**: Python 3.12.0
- **Framework**: Django
- **Platform**: Visual Studio Code (Preferable)

HARDWARE REQUIREMENTS:

- System : MINIMUM i3 and above
- Hard Disk : 40 GB. (min)
- Ram : 4 GB. (min)

key features like brand, model, mileage, fuel type, and year, and leveraging supervised learning models—including Linear Regression, Decision Trees, Random Forests, and SVM—the system delivers precise price estimates while promoting market transparency. Advanced techniques such as RFE, PCA, and ensemble methods like Gradient Boosting further enhance performance, while tools like SHAP and LIME improve interpretability. With strong evaluation metrics and insightful EDA, this scalable, data-driven approach holds immense value for buyers, sellers, dealerships, and online platforms seeking to streamline vehicle valuation through real-time analytics.

Future Scope

While the current model performs satisfactorily, there are several avenues to enhance its effectiveness, usability, and real-world relevance. Expanding the dataset with diverse, international listings and incorporating seasonal and regional trends would boost generalizability. Integrating deep learning models like CNNs for image analysis and LSTMs for temporal forecasting could enrich feature representation. Real-time deployment via APIs and user-friendly web or mobile interfaces would enable dynamic pricing and improve accessibility. Personalization based on factors like urgency or accident history would add practical value, while implementing bias mitigation techniques ensures fairness. Lastly, incorporating user feedback for continuous learning would

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

make the system smarter and more adaptive over time, paving the way for a robust, intelligent car pricing platform.

RESULT

Predicting used car prices is a common application of machine learning, specifically under the domain of supervised learning and regression. In this problem, the goal is to estimate the market price of a used car based on a variety of features such as the make, model, year of manufacture, mileage, fuel type, transmission, and more. Traditional methods of pricing rely on manual inspection or static rule-based systems.

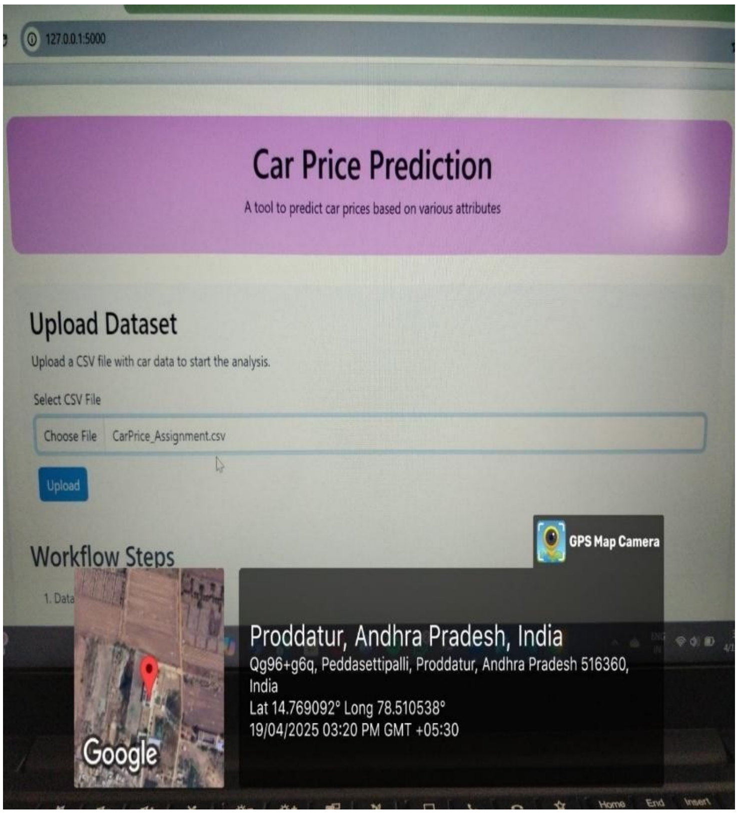

Fig1: Uploading a Dataset

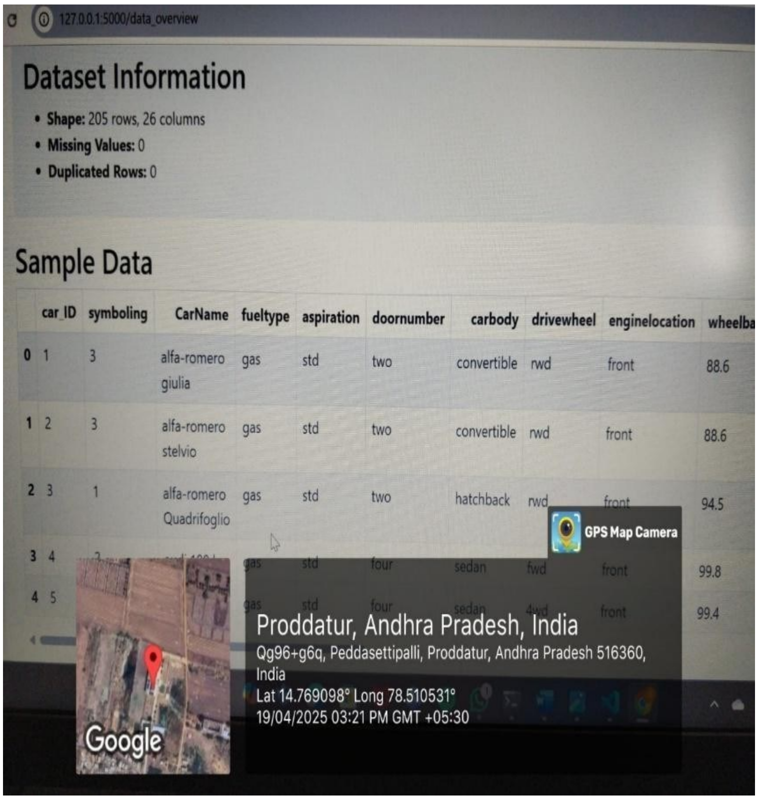

Fig2:Sample Data

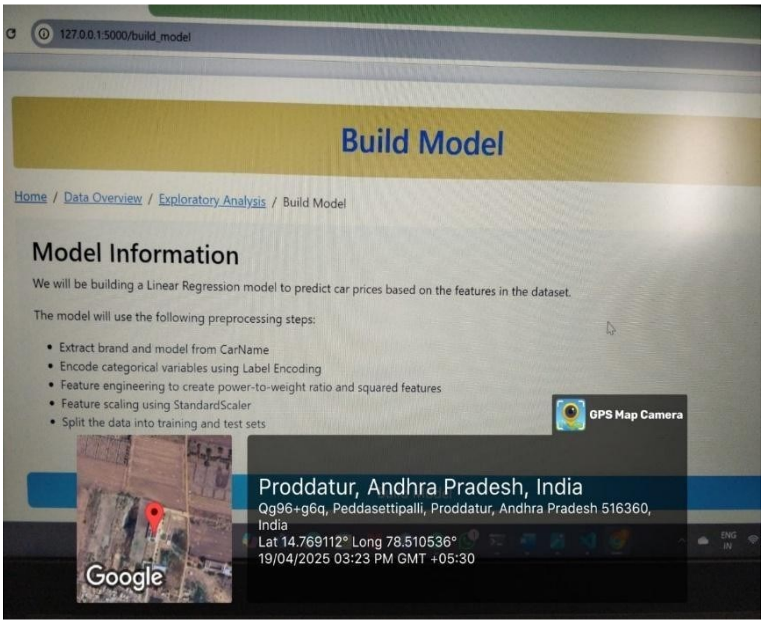

Fig3: Building a Model

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

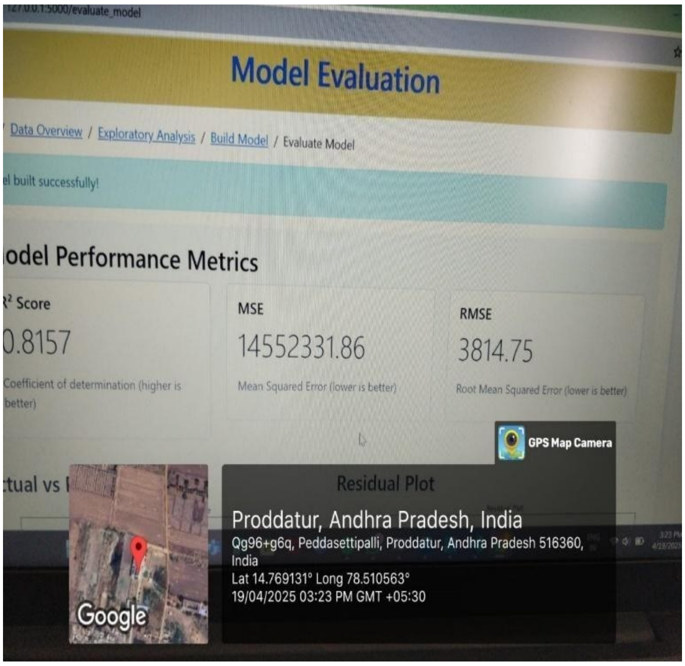

Fig4:Model Evaluation

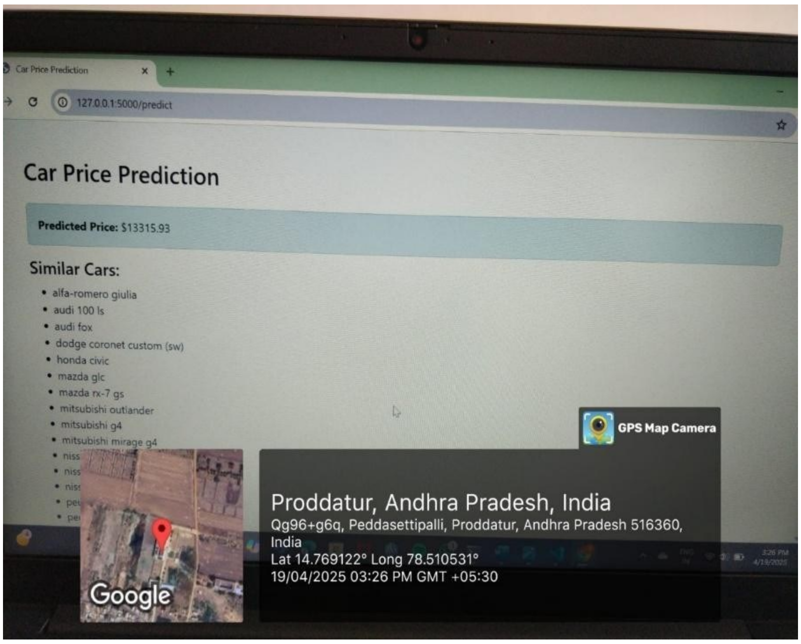

Fig5:Predicted Approx price from the given values ex-1

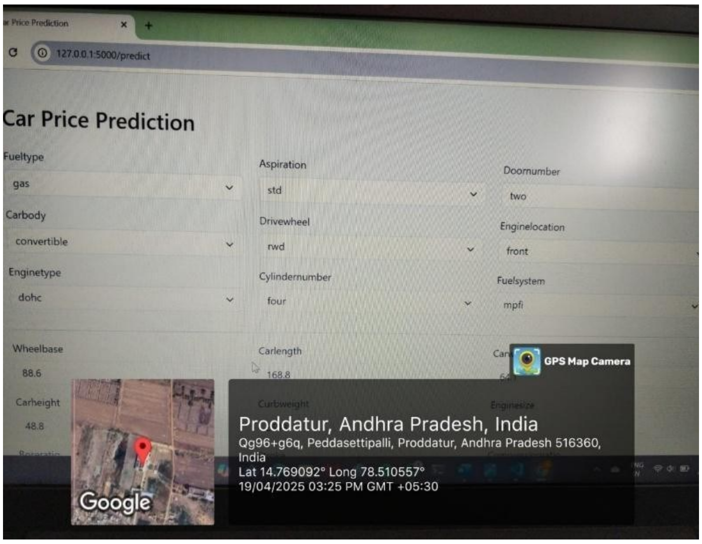

Fig5:Car Price Prediction

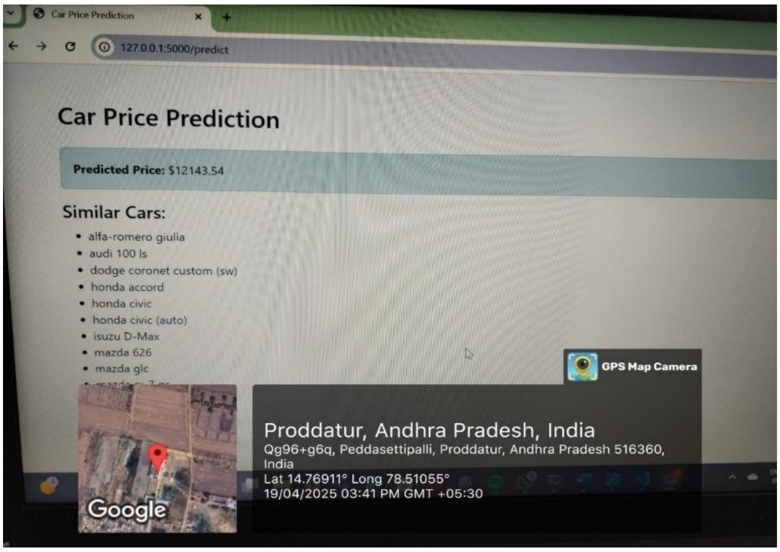

Fig6:Predicted Approx price from the given values ex-2

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

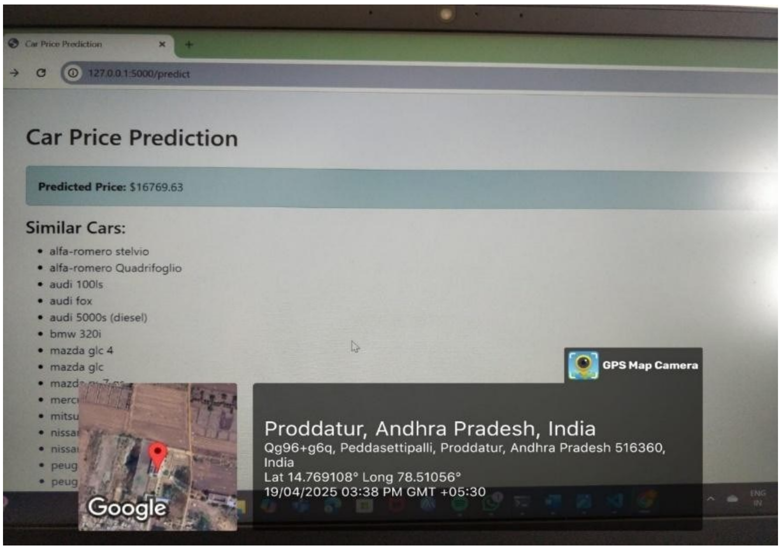

Fig7:Predicted Approx price from the given values ex-2

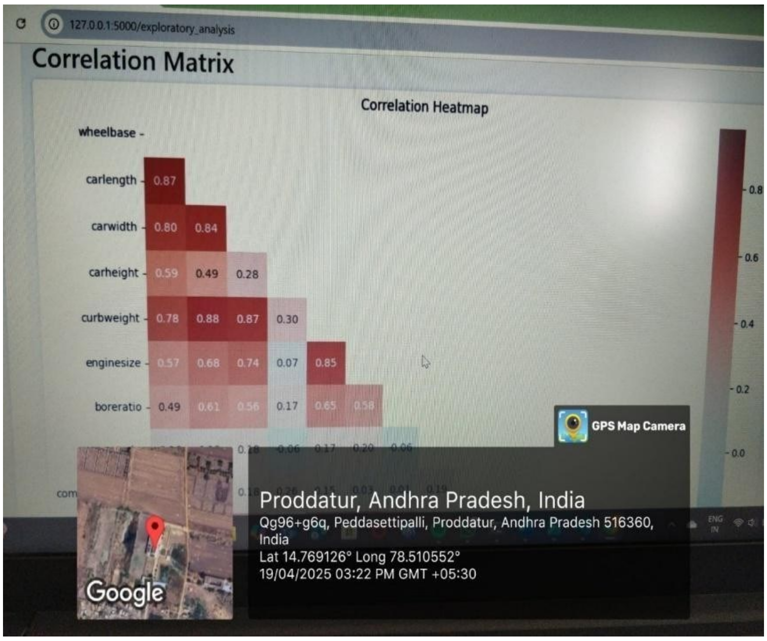

Fig8:Correlation Heatmap

APPLICATIONS

1. **Online Car Marketplaces (e.g., CarGurus, OLX Autos)**

• Provide real-time price estimates to help buyers and sellers make informed decisions.
• Increase trust by flagging "good deals" or "overpriced listings."
• Improve user experience and platform reliability.

2. **Car Dealerships & Resale Businesses**

• Automate trade-in valuations and purchasing decisions.
• Set competitive selling prices based on market trends.
• Enhance profit margins by reducing over- or under-pricing.

3. **Automotive Finance & Insurance**

• Assess car value for calculating loans, EMIs, and insurance premiums.
• Reduce financial risk through consistent and data-backed pricing.
• Avoid human errors and increase fairness in valuations.

4. **Fleet Management & Rental Companies**

• Track car depreciation and predict resale values over time.
• Plan optimal timing for selling or replacing vehicles.
• Improve asset management and cost control.

5. **Individual Buyers and Sellers**

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

• Help private sellers set fair prices for their used cars.
• Assist buyers in evaluating if a listed price is reasonable.

6. Auction Platforms

- Car auction platforms use ML to estimate starting bids and reserve prices based on historical sales and market trends.
- Helps prevent underpricing or overvaluing cars during bidding.

7. Trade-In Estimation Tools

- Many automotive websites offer trade-in calculators powered by ML to give users a fair estimate before visiting a dealership.
- This improves transparency and customer satisfaction.

8. Vehicle Subscription Services

- Services offering cars on subscription use ML to determine pricing tiers based on depreciation, usage, and car model.
- Ensures fair pricing for both provider and subscriber.

9. Car Comparison Tools

- Some websites and apps allow users to compare car models by value and pricing predictions.
- Helps users choose the best-value vehicle within their budget.

10. Supply Chain & Inventory Management

- Dealers and used car platforms use ML to predict the resale value of cars in stock and optimize inventory rotation.
- Prevents overstocking low-demand vehicles or holding depreciating assets too long.

ADVANTAGES

High Accuracy (with the right data):

ML models can capture complex patterns and relationships (like mileage vs. year vs. brand) that traditional methods might miss.

Automation & Speed:

Once trained, models can predict prices for thousands of cars instantly, helping dealerships, platforms, and consumers.

Scalability:

The same model can be adapted across regions or markets with minimal adjustments if relevant data is available.

Data-Driven Decisions:

Helps buyers and sellers make fairer decisions, reducing under- or overpricing based on emotion or guesswork.

Customization:

Models can incorporate user preferences (e.g., fuel efficiency, brand loyalty) to personalize price recommendations.

Cost Reduction

Saves operational costs for businesses by automating valuation processes and reducing the need for expert appraisals.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Improved Customer Trust

Fair and transparent pricing builds trust among buyers and sellers, enhancing brand reputation.

Competitive Advantage

Businesses using ML pricing tools gain an edge over competitors by offering better insights and faster services.

DISADVANTAGES

Data Quality & Availability:

ML models are only as good as the data. Incomplete, outdated, or biased data (e.g. not enough electric cars) can mislead predictions.

Model Complexity:

Some algorithms (like deep learning) are hard to interpret, making it tough to explain why a certain price was predicted.

Market Fluctuations:

Car prices can change fast due to seasonality, economic shifts, or new car launches—ML models might not adapt in real-time unless updated frequently.

Overfitting Risk:

A model might perform well on training data but fail in the real world if it's not properly validated.

Ethical and Fairness Concerns:

If trained on biased data (e.g., favoring certain brands or locations), the model might perpetuate pricing inequalities.

Lack of Transparency

Some models behave like a "black box," making it hard to understand why a certain price was predicted, which may reduce trust.

Sensitivity to Market Changes

Sudden shifts in market trends (e.g., due to economic downturns or fuel price spikes) can make predictions inaccurate until the model adapts.

Bias in Data

If historical data includes biased pricing (e.g., undervaluing certain brands or locations), the model can unintentionally reinforce that bias.

Initial Development Cost

Building a robust ML model involves time, expertise, and resources, which might be a barrier for smaller businesses.

Data Privacy Concerns

Collecting and storing detailed vehicle and owner information must comply with data protection regulations (like GDPR), adding legal complexity.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Conclusion:

This project showcases a powerful machine learning-based solution for accurately predicting used car prices, addressing the critical challenge of price uncertainty in the second-hand vehicle market. By analyzing key features like brand, model, mileage, fuel type, and year, and leveraging supervised learning models—including Linear Regression, Decision Trees, Random Forests, and SVM—the system delivers precise price estimates while promoting market transparency. Advanced techniques such as RFE, PCA, and ensemble methods like Gradient Boosting further enhance performance, while tools like SHAP and LIME improve interpretability. With strong evaluation metrics and insightful EDA, this scalable, data-driven approach holds immense value for buyers, sellers, dealerships, and online platforms seeking to streamline vehicle valuation through real-time analytics.

REFERENCES

1. Agerri, R., & García, F. (2019). Predicting used car prices using machine learning algorithms. Proceedings of the European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), 123-136.
2. Mousavi, M., & Hedayati, M. (2020). Predicting car prices using machine learning algorithms: A comparison of regression models. International Journal of Advanced Computer Science and Applications, 11(4), 243-248.