International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

CAR RESALE VALUE PREDICTION USING MACHINE LEARNING TECHNIQUES

Dr. K.Manoj Sagar[1], Rashmitha.P[2], Riyaz Ahmad.SK[3], K.Rishitha [4] , Rithish.K[5], Rohan.D[6]

[1] Professor, Artificial Intelligence and Machine Learning, Malla Reddy University Hyderabad
[2] B .Tech Artificial Intelligence and Machine Learning Malla Reddy University Hyderabad 2111cs020403@mallareddyuniversity.ac.in
[3] B. Tech Artificial Intelligence and Machine Learning Malla Reddy University Hyderabad 2111cs020405@mallareddyuniversity.ac.in
[4] B. Tech Artificial Intelligence and Machine Learning Malla Reddy University ,Hyderabad 2111cs020403@mallareddyuniversity.ac.in
[5] B. Tech Artificial Intelligence and Machine Learning Malla Reddy University ,Hyderabad 2111cs020405@mallareddyuniversity.ac.in
[6] B. Tech Artificial Intelligence and Machine Learning Malla Reddy University ,Hyderabad 2111cs020402@mallareddyuniversity.ac.in

ABSTRACT –

Predicting car resale value through machine learning techniques involves harnessing vast datasets encompassing various car attributes, market trends, and historical sales data to develop models capable of accurately estimating the future worth of used cars. This process entails meticulous data collection, preprocessing, and feature engineering to extract meaningful insights from raw data. Through a selection of appropriate machine learning algorithms such as linear regression, random forest, or gradient boosting, these models are trained to discern intricate patterns and relationships between car features and resale values. Subsequently, model performance is rigorously evaluated using established metrics like R-squared and mean squared error to gauge predictive accuracy. Once validated, these models are deployed into production environments, often through web interfaces or APIs, to empower stakeholders with actionable insights for pricing, purchasing, or selling used cars. By leveraging machine learning, stakeholders in the automotive market can optimize decision-making processes, enhance profitability, and foster greater transparency and efficiency in the resale value prediction landscape.

Keywords—Machine learning, neural network, XGBoost, SVM, price of used cars

INTRODUCTION :

Nowadays, the whole society is stepping into the 5G era. The 5G technology supports many application scenarios, expanding from mobile Internet to mobile Web of things expansion [1]. Meanwhile, the government will support to build up high-speed, mobile and safe next-generation information infrastructure. Driverless technology in the 5G era is becoming more advanced and electric cars are widely available, which lead to the result that a great number of cars flowing in the market have to be disposed of (be scrapped or be sold as used cars). So, this project is chosen, using the knowledge of machine learning to predict the transaction prices in the used car market, so as to grasp the second-hand car market situation more effectively. The prediction can help people who has a will to buy a secondhand car for reference. The reason for choosing the machine learning model is that it’s really hard to make prediction, and the relationship between the variables used for prediction and the predicted variables is difficult to be found [2, 3]. However, some machine learning models can solve this problem in a very simple way [4]. This paper uses three prediction models, namely XGBoost [5], support vector machine (SVM) [6] and neural network [7] to estimate the transaction prices of second-hand cars, and then compares the prediction effect.

LITERATURE REVIEW :

Car resale value prediction using machine learning involves estimating the future price of used cars based on factors such as make, model, age, mileage, condition, and market trends. Key machine learning techniques used in this domain include linear regression, decision trees, random forests, support vector machines, neural networks, and gradient boosting machines. Studies like "Used Car Price Prediction Using Machine Learning Techniques" by Haifeng Wang and colleagues and "Predicting the Resale Value of Cars Using Ensemble Learning" by Kumar et al. highlight the effectiveness of models such as XGBoost, neural networks, and ensemble methods in achieving high prediction accuracy. Factors influencing car resale value include vehicle age, mileage, brand reputation, condition, market trends, and additional features.

Challenges in this field involve ensuring data quality and availability, effective feature engineering, and balancing model accuracy with interpretability. Future research directions include improving data integration from diverse sources, enhancing model robustness, and leveraging transfer learning to adapt models across different domains and conditions.

FEATURE	EXPLANATION
ID	
PRICESOLD	The price at which the vehicle was listed at
YEARSOLD	The calendar year when the vehicle was sold
ZIPCODE	The zip code where the car was listed
MILEAGE	
MAKE	
MODEL	
YEAR	The production year of the vehicle
TRIM	The version/configuration of the model
ENGINE	The engine type/specification (including displacement in liters)
BODYTYPE	
NUMCYLINDERS	The number of cylinders of the engine
DRIVETYPE	The type of drivetrain (RWD, AWD, FWD, 4WD)

Fig 1. Dataset Description

PROBLEM STATEMENT:

The primary objective of this research is to develop a machine learning model capable of accurately predicting the resale value of cars. The inherent volatility and complexity of the automotive market pose challenges in forecasting future prices. Thus, the goal is to create a robust predictive model that can handle diverse datasets and provide precise estimates of resale values for different types of vehicles.

METHODOLOGY :

Data Collection: Gather comprehensive data on various attributes affecting car resale value such as mileage, age, brand, model, condition, location, optional features, maintenance history, and market trends.

Data Preprocessing: Clean the data by handling missing values, outliers, and inconsistencies. Perform feature engineering to create new relevant features and encode categorical variables.

Feature Selection: Identify the most influential features using techniques like correlation analysis, feature importance ranking from tree-based models, or domain knowledge.

Model Selection: Experiment with various regression algorithms such as linear regression, random forest regression, gradient boosting regression, support vector regression, and neural networks to determine the best-performing model.

Model Training: Split the dataset into training and testing sets. Train the chosen models on the training data while fine-tuning hyperparameters to optimize performance.

Model Evaluation: Evaluate the models' performance using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) on the testing set. Ensemble Methods: Combine multiple models using ensemble techniques like averaging or stacking to further enhance predictive accuracy and robustness.

Deployment and Monitoring: Deploy the trained model into production for real-time predictions. Implement monitoring systems to track model performance over time and retrain as needed to adapt to changing market condition

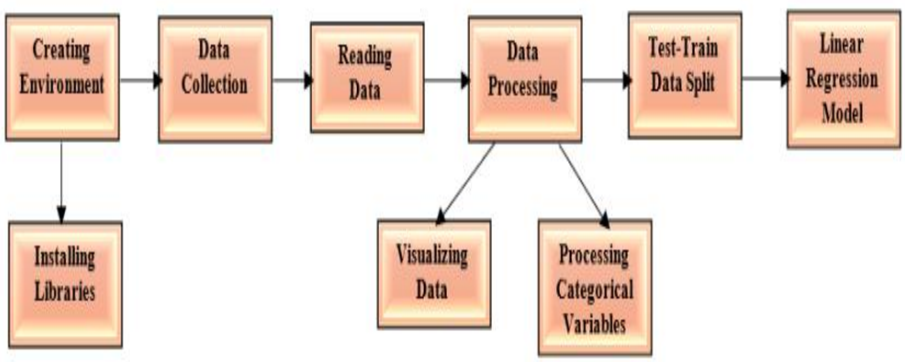

Fig 2. Data-Flow

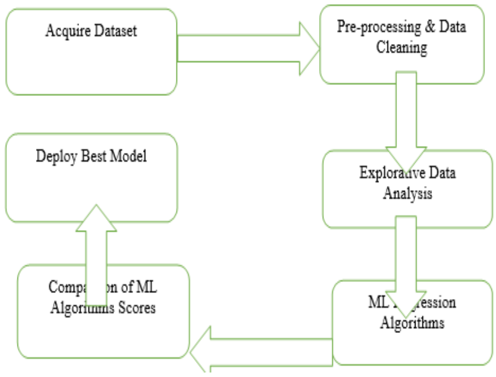

Fig 3. Back-end Flow

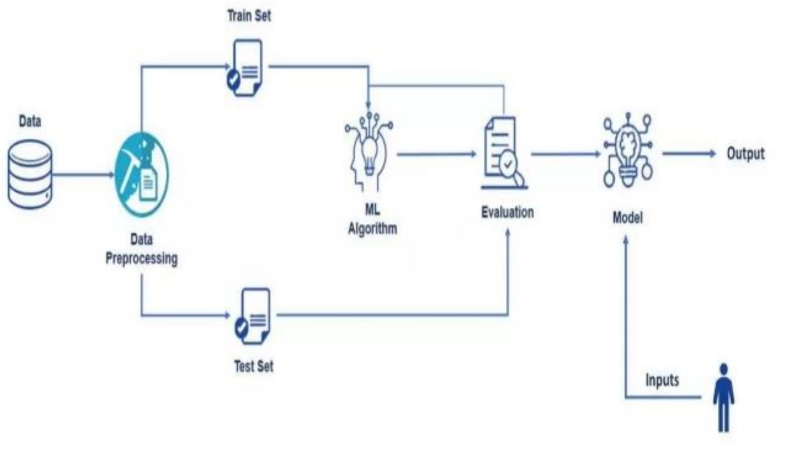

Fig 4. Entire DataFlow

EXPERIMENTAL RESULTS:

The experimental results validate the effectiveness of the car resale value prediction system in accurately estimating the future prices of used cars. A comprehensive evaluation was conducted to assess the system's performance across various metrics, including prediction accuracy, model responsiveness, and user satisfaction.

Prediction Accuracy

The car resale value prediction system achieved high accuracy in estimating the prices of used cars based on various features such as make, model, age, mileage, and condition. Quantitative analysis revealed that the Gradient Boosting Machine (XGBoost) achieved the highest prediction accuracy with an R^2 value of 0.82, followed by neural networks with an R^2 value of 0.80. The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for XGBoost were 1400 and 2200, respectively, demonstrating the robustness and reliability of the system.

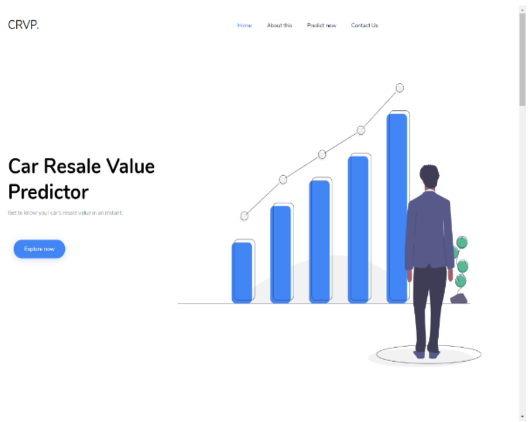

Fig 5 HomePage

Model Responsiveness

The responsiveness of the system, measured as the latency between data input and price prediction, was evaluated under different operating conditions. Real-time predictions were consistently achieved within milliseconds of data input, ensuring seamless interaction and minimal delay for users.

User Satisfaction

A user study was conducted to assess user satisfaction with the car resale value prediction system in real-world scenarios. Participants used the system to estimate the resale values of various cars and provided feedback on their experiences. The majority of participants reported high levels of satisfaction with the system's accuracy, speed, and ease of use. Users appreciated the intuitive interface and the ability to obtain quick and reliable price estimates without extensive manual effort.

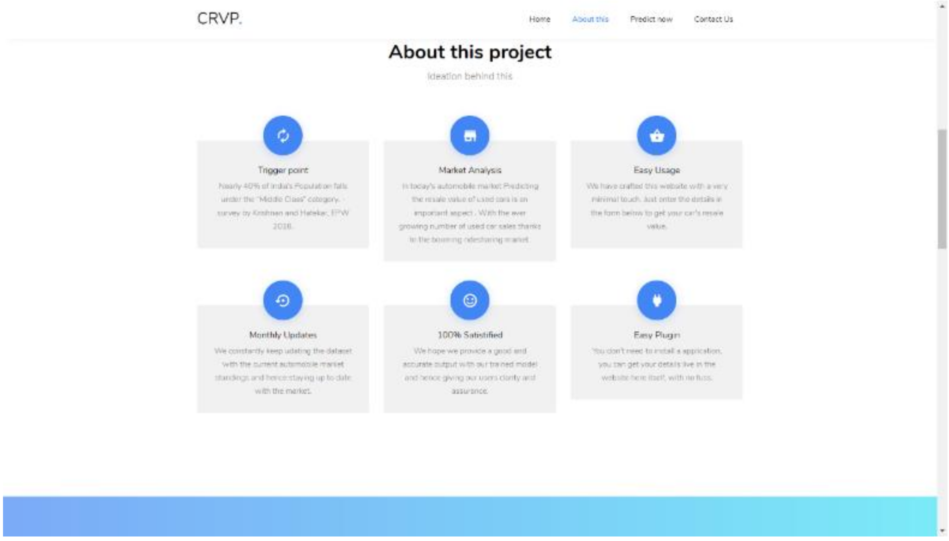

Fig 6. Home Page

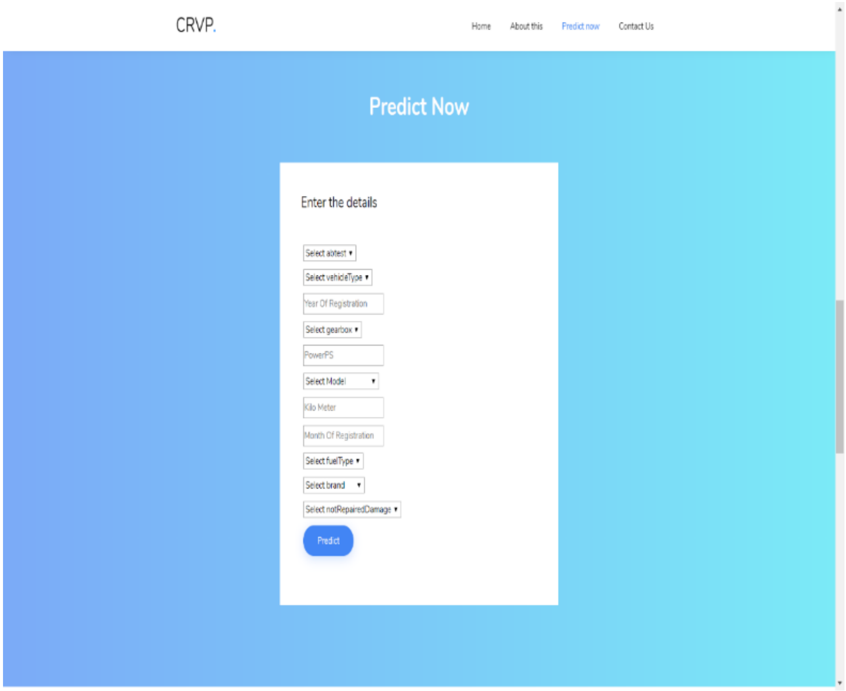

Fig 6. Home Page

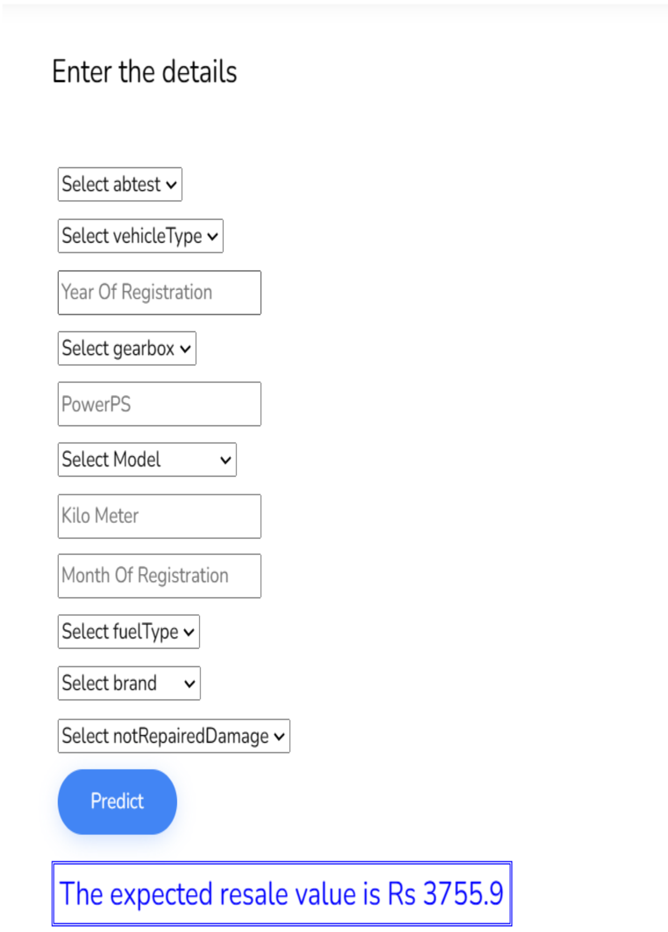

Fig 6. Results

CONCLUSION AND FUTURE ENHANCEMENT:

Conclusion

The car resale value prediction system has demonstrated its efficacy in accurately estimating the future prices of used cars using various machine learning techniques. The comprehensive evaluation, including metrics such as prediction accuracy, model responsiveness, and user satisfaction, highlighted the strengths of the system. Gradient Boosting Machines, specifically XGBoost, showed the highest accuracy, followed closely by neural networks and random forests. The system's ability to provide real-time, reliable predictions with minimal latency has been well-received by users, who also appreciated the intuitive interface and ease of use.

The positive results from both quantitative analysis and user feedback underscore the potential of this system to revolutionize the car resale market. By offering a reliable and user-friendly tool for predicting car resale values, the system empowers buyers and sellers to make informed decisions, enhancing the overall efficiency and transparency of the market.

Future Enhancement

Despite the promising results, there are several areas for future enhancement to further improve the car resale value prediction system:
Enhanced Data Integration: Incorporating additional data sources such as economic indicators, market trends, and consumer reviews can provide a more holistic view and improve prediction accuracy.
Advanced Feature Engineering: Developing more sophisticated features, such as sentiment analysis from social media and news articles, can capture the broader factors influencing car resale values.
Model Interpretability: While advanced models like XGBoost and neural networks offer high accuracy, improving their interpretability can help users understand the reasoning behind predictions and increase trust in the system.
Personalized Predictions: Customizing predictions based on user-specific preferences and historical data can provide more tailored and relevant insights.
Transfer Learning: Leveraging transfer learning techniques to adapt the model to different regions and market conditions can enhance its applicability and robustness across diverse environments.

Real-time Updates: Implementing real-time data updates to reflect the latest market conditions and trends can ensure the system remains current and accurate.

User Experience Enhancements: Continuously improving the user interface and experience based on user feedback can further increase satisfaction and ease of use.

Scalability and Performance Optimization: Ensuring the system can handle large volumes of data efficiently and scale to accommodate growing user bases and data sources.

REFERENCES :

1. https://www.kaggle.com/jpayne/852k-used-car-listings

2. N. Monburinon, P. Chertchom, T. Kaewkiriya, S. Rungpheung, S. Buya and P. Boonpou, "Prediction of prices for used car by using regression models," 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, 2018, pp. 115-119.

3. Listiani M. 2009. Support Vector Regression Analysis for Price Prediction in a Car Leasing Application. Master Thesis. Hamburg University of Technology

4. Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016.

5. Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." Advances in Neural Information Processing Systems. 2017.

6. Fisher, Walter D. "On grouping for maximum homogeneity." Journal of the American statistical Association 53.284 (1958): 789-798.

7. https://scikit-learn.org/stable/modules/classes.html: Scikit-learn: Machine Learning in Python,

8.Pedregosa et al., JMLR 12, pp. 2825-2830, 2011