IT Project

**Used Car Price Prediction Using Machine Learning**

By

Abhiram Pasaladi

Submitted to:

Professor Anshuman Rangaraj

## **Abstract**

Car Resale is a common practice happening in today’s car market interests. It benefits both the sellers and buyers, who want to sell it at a proper price based on usage and also, makes it easier for buyers to afford the used cars and meet their car drive goals. Determining used car prices is difficult because of many factors like odometer reading, car condition, year, make, market condition and demand which affect the prices vastly. For sellers and buyers, it becomes difficult to estimate the prices of resale cars. To make it easier for sellers and buyers, a car price prediction machine learning model is designed in this research. This research explores some machine learning models which can help in accurate price prediction of resale cars based on many features. Although there are many research models to predict car prices, this research covers vast car models, important features and records to make accurate predictions. In this research, a huge dataset consisting of records of many features and car models dating through some years is considered. This research mainly focuses on utilizing this huge dataset for accurate price prediction containing many features and records. Many models were used and compared to find the right model for prediction. Two of them came out with near accurate predictions, namely Random Forest regression and Stacked Regressor. To achieve this, many dataset handling techniques like data cleaning, preprocessing, removing outliers etc... were considered, which were discussed later in this paper. To facilitate car price prediction for sellers, a web app was also designed using Radom Forest model giving near accurate predictions rating to 82 percent. Although the model has successfully achieved near accurate prediction using huge dataset with many car models, it has come up with some limitations as well, opening research for further increase in prediction scope and accuracy along with increase in considering new features.

**Keywords:** Car Price Prediction, Machine Learning, Used Cars, Regression, Used Cars, Cars Resale, Ensemble Models

## **Table of Contents**

[Abstract](#_Abstract) .......................................................................................................................................... 2

[1. Introduction ................................................................................................................................](#_1._Introduction) 6

[1.1 Background and Research context ………………………………………………….](#_1.1_Background_and)... 6

[1.2 Problem Statement ………………………………………………………………….](#_1.2_Problem_or)... 6

[1.3 Project Objectives ………………..………………………………………………….](#_1.3_Project_Objectives).. 6

[1.4 Significance and Motivation ……………………………………………………….](#_1.4_Significance_and).... 7

[1.5 Literature review and Gaps in Literature …………………………………………](#_1.5_Literature_review)….. 7

[1.6 Methodology and approach ………………………………………………………...](#_1.6_Methodology_and).... 8

[1.7 Structure of the paper ……………………………………………………………….](#_1.7_Structure_of)..8  
  
2. [Literature Review](#_2._Literature_Review) ....................................................................................................................... 8

[2.1 Introduction to the topic …………………………………………………………….](#_2.1_Introduction_to)... 9

[2.1.1 Regression-based Approaches ………………………………………..…](#_2.2_Historical_context)…. 9

[2.1.2 Tree-Based and Ensemble Approaches ……………………………….…...](#_2.3_Theoretical_Framework) 9

[2.1.3 Deep Learning Approaches ……………………………………………...](#_2.4_Methodologies_and)... 9

[2.2 Data Sources and Feature Engineering………………………………………](#_2.5_Key_Concepts)……... 10

[2.3 Gaps in Existing Research and Future Scope…………………………….](#_2.10_Research_trends)........... 10

[2.4 Conclusion ……………………….…………………………………………](#_2.13_Conclusion_and)............ 10

3. [Methodology](#_3._Methodology) ............................................................................................................................. 11

3.1 [Dataset Acquisition](#_3.1_Data_Acquisition) ..................................................................................................... 11

3.2 [Data Preprocessing](#_3.2_Data_Preprocessing) ..................................................................................................... 11

3.3 [Data Visualization](#_3.3_Data_Visualization) ...................................................................................................... 12

3.4 [Data Modelling](#_3.4_Data_Modelling) ........................................................................................................... 14

3.5 [Performance Evaluation](#_3.5_Performance_Evaluation) .............................................................................................. 14

3.6 Web Application for Prediction …………………………………………………….. 15

4. [Results and Discussion](#_4._Results_and) .............................................................................................................. 16

5. [Conclusion and Future Work](#_5._Conclusion_and) ................................................................................................... 17

[5.1 Research Conclusion…..……………………………………….………………..](#_5.1_Limitations)..... 17

5.2 Future Work ……………………………………………………………….……….. 18

[References](#_References) ................................................................................................................................... 18

## **List of Used Figures**

Figure-1: Data Types for features in Dataset …………………………………………………. 12

Figure 2: Outliers detection in price and Odometer features of Dataset ……………………... 12

Figure 3: Features Correlation Analysis ……………………………………………………… 13

Figure 4: Feature Importance Bar diagram …………………………………………………… 14

Figure 5: Web Application to predict Car prices ……………………………………………... 15

## **List of Tables**

Table 1: Results of Different Machine Learning Models with metric scores …………………. 16

## **1. Introduction**

**1.1 Background and Research Context**

The demand for used cars in the market has increased a lot over the past decade, forced by some factors like the high price of new cars, increased vehicle age, and individual’s economic uncertainty [1]. Buyers mostly seek for used cars to reduce costs and get a proper running vehicle. However, finding the correct price of a used car is difficult which depends on many features like odometer reading, age, manufacturer, condition etc. [2].

Traditional pricing for used cars depends on manual assessments, where third-party experts or dealers estimate the prices based on past sales and used car measurements [3]. This process lacks in consistency and also fails to consider the ever-changing market fluctuations. Here, machine learning helps in providing this solution using data driven approach using past and huge data, to help in making accurate predictions [4].

Machine learning techniques like decision trees, ensemble models and deep learning architectures are being used to improve the price estimation accuracy [5]. By using the historical sales dataset with many features and records, predictive models can automate and estimate the price prediction for both the buyers and sellers [6].

**1.2 Problem Statement**

Estimating the price of used cars for resale accurately is a problematic issue due to many features and factors. Buyers and Sellers both depend on online listings, dealer decisions or some third-party valuation tools that may not correctly estimate the real-time market [8]. Manual methods might lead to subjective and inconsistency, often leading to overpricing or underpricing issues [9].

The main problem this research addresses is the use of huge datasets consisting of many car models and records and overcoming the research issues which lack an automated and highly accurate car pricing models considering many manufacturers and car types. Existing models have made considerable progress in improving price prediction, but also, they usually fail to consider huge dataset consideration for many car model records along with overfitting, data discrepancies and feature selection issues [10]. This research study goal is to implement machine learning models that help in predicting car prices for multiple car models by comparing different metrics and also help the buyers and sellers in predicting car prices using a web application for car predictions. This web application provides the users with so many manufacturers of cars providing prediction as per their car requirements liking.

**1.3 Project Objectives**

This research focuses on fulfilling the following objectives:

* To create a ML model that accurately helps in estimating prices of used cars using a huge past cars models data and features [1]. Also, make this model usable to create a car price predictor to help buyers and sellers.
* To compare different regression techniques and find the best model providing good accuracy [2].
* To find out key features affecting the car prices and measure their importance using feature selection methods [3].
* To validate the effectiveness of the models using following metrics like Mean Absolute Error, R2 score also called Rsquared, RMSE – Root Mean Squarred Error and MSE - Mean Squared Error [5].

**1.4 Significance and Motivation**

Predicting the used cars prices accurately is important for buyers, sellers, dealers, insurance companies and financial institutions [6]. Fairness in transactions, reduction of overpricing or underpricing, and increased buyer-seller trust are all provided by an objective valuation technique [7].

Also, additionally, predictive analysis can help dealers optimize their inventory pricing techniques by predicting future demand signatures [8]. Lenders and insurance companies can also use these predictive models to rightly evaluate vehicle prices for actions like loan approvals, leasing plans, and coverage policies [9].

With the increase of digital automotive marker, integrating AI-based pricing tools in platforms like cars24 and TrueCar has further improved the importance of machine learning in estimating vehicle price evaluation [10]. This research study contributes to the growing field by using large datasets with many car models and measuring utilizing machine learning to accurately estimate car prices.

Additionally, a python Flask web application designed through this research machine learning models helps in improving practical utility and reaching a wider range of people. This research gives a user-friendly interface which can predict the car prices accurately.

**1.5 Literature Review and Gaps in Literature**

While several studies have gone through many machine learning approaches for predicting car prices, many existing models have been found with limitations in handling various diverse data sources, missing values in records, and regional changes in market [1]. Some studies have some limitations like using small datasets, considering only limited car company records or with only limited number of features. So, there is a need for models that consider a comprehensive set of car related features and are trained considering large datasets. Some methods face issue with overfit in training data due to less feature selections, which leads to poor generalization of records, on unseen datasets [2].

Also, deep learning techniques, like Artificial Neurral Networks and Long Short Tem Memory networks techniques, are being less utilized in prices predictions due to their high requirements in computing [3]. This research study covers these gaps by using huge datasets with various company records and data along with a hybrid approach that uses both the traditional regression techniques with advanced ensemble learning models, optimizing performance through feature engineering [4]. Other previous research works fail to consider many different car models and solely focus on prediction accuracy diminishing the consideration of various car models, while my research considers both to make accurate predictions. This research provides a web application to predict car prices of different kinds providing an opportunity for both the buyers and sellers to use.

**1.6 Methodology and Approach**

This research follows a structured methodology to create a price prediction model, where it strats with the data collection from online data source Kaggle data source. The huge dataset with various records undergoes many steps like preprocessing steps, including data cleaning, handling missing values, removing outliers in datasets, performing standardization and encoding categorical variables [5].

Many machine learning models were implemented, and trained with data preprocessed dataset containing historical car price data. These models are then evaluated using statistical metrics, and next their performance metric values are compared to determine the most accurate machine learning prediction model [6]. Additionally, feature importance analysis, outlier analysis, correlation analysis are also made to identify important factors effecting prediction of car prices [7].

**1.7 Structure of the Paper**

The research study paper is structured into 5 main parts as shown below:

1. **Introduction** – which provides background and context, problem statement, project objectives, significance and motivation, Literature review briefly and methodology overview.
2. **Literature Review** – which talks on prior research, theoretical models, existing methodologies, and identified research gaps.
3. **Methodology** – which clearly provides details on data acquisition, preprocessing, model selection, and evaluation techniques.
4. **Results and Findings** – which showcases findings, model performance comparison, and analysis of prediction accuracy.
5. **Conclusion and Future Work** – which summarizes key contributions, limitations, and provides scope for future research.

**2.** [**Literature Review**](#_2._Literature_Review)

The estimation of prices for used cars has always been a big challenge, as it gets influenced by many factors like age, odometer reading, brand, model, fuel type, drive type and overall market conditions. Traditionally, pricing for used cars was decided by expert valuations and manual assessments, which had issues with the subjectivity concept and inconsistencies. With the improvement in AI and data analytics, machine learning became one of major powerful automated tools to automate and also to increase the accuracy while predicting the prices of used cars. Several research studies have provided and explored many machine learning techniques which range from some regression models to ensemble learning and then more advanced deep learning techniques, which can help in improving accuracy of prediction in used car prices. This section of research paper provides with a summarized review of literatures already existing, showcasing key methodologies, trends, and gaps in research studies which are related to machine learning-based used car price prediction.

**2.1 Machine Learning usage in Used Car Price Prediction**

Utilization of Machine Learniing (ML) Techniques in predicting the car prices for resale have gained major popularity in recent years. Many algorithms like decision trees, artificial neurral networks (ANNs), and , support vector machines (SVMs), ensemble models like XGBoost and Random Forest, have mostly been used for research and study. Research has shown that ensemble learning techniques perform well and show accuracy better than traditional linear regression models (work well on linear data) and other models handle better nonlinear relationships in data [1].

**2.1.1 Regression-Based Approaches**

Regression models are very mostly used in case of predicting continuous variables scenarios, making them a basic fundamental approach in estimating car prices. Studies have used and measured the effectiveness of Linear Regression, Ridge Regression, and Lasso Regression, which gave an understanding that these models give a baseline for price prediction, but they struggle and are inaccurate because of feature interactions and nonlinearity in data [2]. Research which was comparing XGBoost and AdaBoost regressors has shown that these boosting techniques greatly improve predictive accuracy by decreasing data bias and variance compared to linear models [3].

**2.1.2 Tree-Based and Ensemble Models**

Tree-based models, namely Random Forest and Decision Trees, were mostly used in automotive price prediction. Particularly, Random Forest algorithm has shown great predictive performance as compared to other models due to the algorithm’s ability to decrease overfitting [4]. LightGBM and XGBoost were chosen in some research studies for their ability to handle computational efficiency and ability to handle large datasets [5]. Comparative analyses in research showed that ensemble methods perform better than single regression models by considering and capturing complex relationships present between features [6].

**2.1.3 Deep Learning Models**

ANNs (Artificial Neurral Networks) and deep learning architecturres such as Long Short-Tem Memory (LSTM) and Convoltional Neurral Networks (CNNs) networks were also considered in research to predict the prices of resale cars. Some research studies also show that ANNs can consider complicated patterns present in the pricing data, but they depend on and require very extensive tuning and also large datasets requirement to perform well [7]. Using ANN models also showed some competitive results, but they are pricier in computational activities compared to other tree-based methods [8].

**2.2 Data Sources and Feature Engineering**

An important aspect while using machine learning based prediction for car prices is the consideration of selection and preprocessing of features. Many research studies gather data from online automotive marketplaces, which are large-scale and also real world past transactional data records [9]. Feature engineering techniques like the one-hot encoding technique used for categorical to numerical variables convertion, including standardization of numerical variables, and also outlier removal, help greatly in improving performance of a model. Research Studies showcase on how important feature selection processes are, like Recurrsive Feature Elimination (RFE) and SHAP (SHapley Additive exPlanations), which help in finding out the most impactful variables affecting the prices of cars in prediction [10].

**2.3 Gaps in Existing Research and Future Directions**

While existing studies have greatly provided a contribution to improve prediction in case of predicting car prices, many research gaps remain. Firstly, the use of generalized and huge dataset considering many car models is one of the research gap. Also, many models used in this research lack generalizability standards across many different regions due to differences and variations in dynamics of market. The previous works solely focus on accuracy predictions rather than many car models of dataset during training, where my Research covers this gap by considering the both to produce near accurate predictions. Also, this research is unique in data cleaning, where it focuses both on prediction accuracy and different car models consideration for training, without over diminishing the dataset just for prediction. This research covers the gap of utilizing the ml models and data featuring techniques to create an application to predict prices of many different car models. In addition to that, some external economic factors like inflation, fuel price changes, and seasonal trends are also not often considered and are not integrated into the models design for prediction of prices. Future research studies should also focus on considering and adding in real-time data streams, macroeconomic indicators, and hybrid modeling approaches which integrate machine learning with econometric models. Also to add, explainable AI (XAI) techniques can also be used to improve model understanding and transparency for all the stakeholders like sellers, buyers, dealers etc.

**2.4 Conclusion**

Using Machine Learning to estimate prices of cars has clearly shown as to be very valuable compared to traditional pricing models. Ensemble techniques, particularly Random Forest, XGBoost, and LightGBM, have clearly shown good predictive accuracy as compared to the other models. However, deep learning models are being used less due to their high computational requirements. Future research have the scope in improving model robustness, integrating external economic factors, and improving model interpretability to make sure valuations are fair and also reliable.

## **3. Methodology:**

The methodology used here is to create a price prediction model which can accurately predict and estimate the car prices for the resale. The huge dataset is open source and is taken from Kaggle. It has many features and records necessary to predict car prices, thus being an ideal data source to make predictions. The main goal is to understand the dataset and make necessary steps like cleaning dataset, preprocessing, feature engineering, data standardization and building a model to estimate car prices accurately.

**3.1 Dataset Acquisition**

The dataset for this project is taken from a openly available source Kaggle. It had a total of 5,25,839 records along with 22 columns containing many car features like vehicle specifications, pricing, odometer readings, manufacturer, year, paint color, condition, fuel, drive, cylinders etc. But initially dataset had discrepancies like having missing values, inconsistencies in categorical column fields and some outliers in columns with numerical values.

**3.2 Data Preprocessing**

In this data preprocessing stage, many steps were considered to clean the dataset and make it workable for effective ML models prediction. Initially, many unnecessary columns that did not help in modelling like city, url, long, lat, make etc… were removed to reduce the huge dataset. Next, to solve the problem of missing values, whichever records (rows) with incomplete or empty data were removed. Outliers were also removed by considering extreme price values like removing prices under $1000 and above $100000, along with odometer readings which were exceeding 3,000,000 miles so as to make it realistic with the dataset. Finally, after data cleaning, dataset consisted of 1,03,045 rows and 11 columns. Categorical data like manufacturer, fuel type, transmission, color, drive type, cylinders etc. were standardized by making them to lowercase and encoding using the Label encoder to support in Machine learning preprocessing. Also, a new feature “car age” was created after subtracting between current year 2025 and the vehicle’s manufacturing year to improve prediction accuracy. Next, the dataset which was cleaned, was next trained with 80% and tested with 20% dataset splitting to properly measure performance of model. At last, numerical columns were standardized using standardscaler to make sure data distribution is uniform.

**3.3 Data Visualization**

Data visualization is a fundamental tool which helps in exploring and understanding datasets. This also helps us in performing data preprocessing steps and data cleaning steps to identify outlier patterns which might impact the prediction. After cleaning the dataset, the final dataset has 1,03,045 records with 11 columns. The data frame has 8 categorical (object type**:** manufacturer, condition, cylinders, fuel, transmission, drive, type, and paint\_color) and 3 numerical features (odometer, car\_age of float64 type, price of int type). The dataset’s memory usage is at 9.4 MB, showing optimized structure.

A screenshot of a computer code

AI-generated content may be incorrect.

**Figure-1: Data Types of features in Dataset**

Outliers of the dataset are identified using the bar plots on numerical features price, odometer and car\_age. Many records were more than $40,000 and only with some near to $1,00,000 limit. Here, prices of more than $100000 were removed instead of $40,000 as the records decrease and manufacturers consideration decreases impacting input features from user. Also, odometer readings in bar plot showed outliers exceeding 2,500,000 miles, where filtering is done are 3,000,000 miles for this project to consider more records, which impact linear regression.

A comparison of a graph

AI-generated content may be incorrect.

**Figure 2: Outliers detection in price and Odometer features of Dataset**

To clearly understand the relationships existing between the features of cleaned dataset and the car\_price target variable, a correlation heatmap was created. This heatmap showcased some correlations, mainly negative correlation (-0.35) is seen between car\_age and price, showing that older cars have lower prices. Also, negative correlation (-0.34) was observed between odometer and price, supporting the higher mileage cars have lower price rate in market. Interestingly, cylinders feature was observed with positive correlation (0.24), suggesting that vehicles which have more cylinders were tagged with high car\_price, may be due to performance factors or luxury branding manufacturer.

A blue and red squares with white text

AI-generated content may be incorrect.

**Figure 3: Features Correlation Analysis**

A feature importance plot was also created using Random Forest regressor to understand price dependency on price rate consideration. The results showed car\_age as the most important feature in price consideration, which was same as the correlation analysis. Unexpectedly cylinders feature was placed above odometer in importance, which realistically was not true. This discrepancy might be because of random forest measured feature importance supporting variables that help with clearer data splits. While cylinders feature sometimes directly effects vehicle classification (eg: economy vs performance cars), this model might have lessened odometer feature importance due to moderate collinearity with feature car\_price. This is showing the limitation of using tree-based feature importance.

A graph with blue bars

AI-generated content may be incorrect.

**Figure 4: Feature Importance Bar diagram**

Overall, the combination of outlier detection, correlation analysis and feature importance have given a comprehensive understanding of dataset and features making the accuracy in prediction more realistic. These data visualization insights helped in data cleaning, preprocessing, model selection, feature engineering and output results.

**3.4 Data Modelling**

To create a reliable model for prediction of vehicle prices, many regression techniques were implemented and compared. The models which were used are Linear regression, Stacked Regression (that combined decision tree, random forrest and gradient boosting), Random Foresst, Gradient Boosting, Decision Tree Regresors, and Support Vector Regressor models. All these ml models were trained on cleaned dataset obtained after preprocessing to correctly make predictions. These models were later evaluated to find out which model gave high performance and accuracy. The stacked regression and random foresst regressor showcased itself with the best performance result outcomes out of remaining models showcasing their abilities to handle nonlinear data.

**3.5 Performance Evaluation**

These models performance were evaluated based on some performance metrics to understand their prediction accuracy. The Mean Absolute Error- MAE metric was utilized in order calculate average magnitude for errors, which provided insights into deviation of predicted values from actual values. The Mean Squared Error metric (MSE) was to quantify the average of the squared discrepancies between actual value and predicted value, addressing large errors. Root Mean Squarred Error metric - RMSE was used to make it easier to compare across different models. The R-squared score was used to find out the variance proportion in prices of vehicles given by models, showing overall reliability of models on prediction. The results obtained showed that ensemble models, mainly Stacked Regression and Random Forest Regressor, produced the best metrics performance between bias and variance, which made them as the final choices for designing a predictive model.

**3.6 Web App for Prediction**

To make use of this prediction model, a web application “Car Price Predictor” was designed, which provides a visually appealing UI. Features like manufacturer, fuel type, transmission, condition, paint color, cylinders were selected using dropdowns, providing easy selection. For numerical features like odometer and year, input text fields were considered. After taking inputs and clicking on Predict button, it gets redirected to new page showing the estimated resale price. This web app workflow provided user friendly, visually appealing and seamless price prediction based on trained models in my machine learning project.

A screenshot of a car price prediction form

AI-generated content may be incorrect.

A blue and white rectangle with text

AI-generated content may be incorrect.

**Figure 5: Web Application to predict Car prices**

**4. Results and Findings**

The machine learning models that were used for modelling provided some important insights into the key factors impacting the price evaluation of resale vehicles. The Random Forest and Stacked Regression models came out with good prediction accuracies, both having the Rsquare values of 0.82. Among all the features, car\_age came out as an important and most influential feature in impacting the prices of resale vehicles, showing high negative correlation with price (-0.35), showcasing the expected decrease of price for older vehicles.

A table with numbers and letters

AI-generated content may be incorrect.

**Table 1: Results of Machine Learning Models with metric scores**

Other models like Linear Regression, SVM showed a very lesser accuracy displaying their inaccuracy with non-linear data. Whereas the other models like Decision tree, Gradient Boosting were showing accuracy near to Random Forest and Stacked Regression but still lesser than them.

Also, in the findings surprisingly cylinders came out as the second most important feature after car\_age, despite having positive correlation with the price (0.24). This observation clearly tells that cars with more cylinders, usually performance or luxury cars, have higher prices. It occupied second place over odometer, which had more negative correlation with price (-0.34). This clearly showcases on how Random Forest’s impurity reduction can override certain features. This proves that the odometer impact was slightly reduced due to collinearity with the car\_age feature (0.22), usually higher mileage indicates lower car price when for resale.

The correlation matrix also provided some findings like negative correlation existing between fuel type and price (-0.23). Also, less impact from feature drive (-0.18) type. Additionally, features like manufacturer and paint color features shoed minimal correlation with price, showing limited predictive power.

Overall, the models implemented captured the key market trends, with Stacked Regression and Random Forest models giving greater consistency and accuracy. But, the results and findings obtained clearly show the need for improvement in feature treatment and handling outliers to improve prediction accuracy.

**5. Conclusion and Future Work**

**5.1 Research Conclusion**

This research study showcased on how effective the ML models can be in accurately predicting used car prices to do resale, with mainly focusing on ensemble ml techniques, Stacked Regression and Random Forest. These models effectively considered and captured non-linear relationships present between the features of cars, also handling the complexities present in large datasets of used car listings. The results and findings showed that ensemble techniques give high accuracy prediction by combining the strengths of multiple models, thereby helping in mitigating overfitting, and also improving generalization across unseen data [1].

One important observation from this research was the challenge that occured with feature importance ranking through Random Foresst ml model. This model, although very useful for non-linear datasets, highlighted features like cylinder count as the second most important feature due to its impurity reduction mechanism, sometimes overriding other relevant feature like odometer reading. This feature importance discrepancy shows how the multicollinearity is impacted in feature importance assessments, mainly among variables like odometer and car\_age. Implementing methods like advanced feature selection, like Permutation Importance, could greatly help in providing a more balanced evaluation of feature importance [2].

While the models successfully covered broad market trends, certain real-world factors, like regional variation in pricing concept, fluctuations in market demand, and vehicle service history, were not incorporated and considered into the dataset used for modelling. This exclusion of such important real world external influences limits the overall capability of models to make accurate predictions. Adding additional dataset features, like regional economic indicators, local dealership pricing trends, and vehicle service histories, could greatly help in improving the accuracy and dependency of price predictions [3].

**5.2 Future Work**

Future work in this research project should focus mostly on improving techniques of feature engineering, by using advanced selection methods like SHAP (Shapley Additive Explanations) and Recursive Feature Elimination (RFE) which can help in improving interpretability and also help in mitigating redundant variables [4]. Also, adding external dataset features, like regional economic indicators, dealership inventory trends, and real-time market demand, would help in giving a good comprehensive approach to estimate the price of resale cars, improving model’s adaptability to ever changing dynamics of market conditions [5]. Another important promising direction can also be like analyzing variations of regional and local pricing by separating datasets based on different geographical factors, which would help models to cover data related to location-specific trends and that locations specific consumer behaviors [6]. Presenting the outliers and data-balancing challenges is very important. For this, using advanced anomaly detection techniques, data-balancing strategies could help greatly in reducing biases which are introduced by rare or extreme cases or features in the dataset records [7]. Furthermore, while this research study has mainly focused on creating a global model performance, the future studies should focus on interpretability techniques at an individual prediction level, thereby allowing for deeper insights into how each feature of cars effects in estimating the prices of resale cars [8]. By reconsidering and improving these areas, future research can help in covering the gap between theoretical modeling methods and real-world applications, which help in designing a more accurate, transparent, and reliable machine-learning-driven price evaluation system for used cars [9].

**References**

1. Vaneesha KH, Srinivas V, Abhishek V, Srinivas S. Comparative analysis of machine learning algorithms for used car price prediction. *Int J Curr Sci Res Rev.* 2024;7(9):Article 39. doi:10.47191/ijcsrr/V7-i9-39.
2. Gupta N, Pareek P, Reis MJCS, eds. *Cognitive Computing and Cyber-Physical Systems: Third EAI International Conference, IC4S 2022, Virtual Event, November 26–27, 2022, Proceedings.* Springer; 2022.
3. Reddy S, Kumar S. A comparative analysis of XGBoost model and AdaBoost regressor for prediction of used car price. *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023).* 2024:441-446. doi:10.5220/0012510700003739.
4. N, Assi A, Shakhshir H, Salman A. Machine learning-powered mobile app for predicting used car prices. *Proceedings of the 2021 3rd International Conference on Big-data Service and Intelligent Computation (BDSIC).* November 19-21, 2021; Xiamen, China. ACM. doi:10.1145/3502300.3502307.
5. Satapathy SK, Vala R, Virpariya S. An automated car price prediction system using effective machine learning techniques. *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES).* 2022:402-408. doi:10.1109/CISES54857.2022.9844350.
6. Yang J, et al. Metaverse: design of the car price prediction model through a machine-learning approach. *2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom).* 2023:734-737. doi:10.1109/MetaCom57706.2023.00139.
7. Hemendiran B, Renjith PN. Predicting the prices of the used cars using machine learning for resale. *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS).* 2023:1-5. doi:10.1109/SCEECS57921.2023.10063133.
8. Wang F, Zhang X, Wang Q. Prediction of used car price based on supervised learning algorithm. *2021 International Conference on Networking, Communications and Information Technology (NetCIT).* 2021:143-147. doi:10.1109/NetCIT54147.2021.00036.
9. Varshitha J, Jahnavi K, Lakshmi C. Prediction of used car prices using artificial neural networks and machine learning. *2022 International Conference on Computer Communication and Informatics (ICCCI).* 2022:1-4. doi:10.1109/ICCCI54379.2022.9740817.
10. Li Y, Li Y, Liu Y. Research on used car price prediction based on random forest and LightGBM. *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA).* 2022:539-543. doi:10.1109/ICDSCA56264.2022.9988116.