Original Scientific Article Journal of Decision Analytics and Intelligent Computing
Vol. 3 issue 1, (2023) 167-184 https://doi.org/10.31181/jdaic10008102023n

Pre-owned car price prediction by employing machine learning techniques

Mauparna Nandan[1,*], and Debolina Ghosh[2]

[1] Department of Computer Applications, Techno Main Saltlake, Kolkata, India

[2] Department of Information Technology, Haldia Institute of Technology, Haldia, India

* Correspondence: mauparna2011@gmail.com

Received 7 May 2023

Accepted for publication 27 September 2023

Published 8 October 2023

Abstract

Pre-owned automobiles including cars are becoming incredibly popular. There has been a steady increase in automobile production namely, passenger cars over the preceding decade with more than 70 million passenger cars being manufactured in 2016 itself. This has given rise to the resale automobile market, which has become a thriving business in its own right. Customers who are interested in purchasing a pre-owned car frequently face the difficulty in locating a vehicle that fits within their financial constraints as well as estimating the price of a specific pre-owned car. Customers can make more educated decisions regarding the purchase of a pre-owned car if they have access to accurate price projections for pre-owned cars. With the proliferation of digital marketplaces, both the buyer and the seller remain more updated regarding the recent market trends and patterns that impact the value of a used car. In this paper, we investigate this issue and propose a forecasting system using machine learning techniques that enables a prospective buyer to anticipate the price of a pre-owned vehicle of interest. The process is conducted with the collection and pre-processing of a dataset followed by an exploratory data analysis. Various machine learning regression techniques, such as Linear Regression, LASSO (Least Absolute Shrinkage and Selection Operator) Regression, Decision Tree, Random Forest, and Extreme Gradient Boosting, have subsequently been implemented. The techniques are then compared so as to determine an optimal solution. Three types of errors namely, MAE, MSE and RMSE have also been calculated in order to determine the best-fitted model.

Keywords: Price Prediction, Machine Learning, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE)

1. Introduction

The pre-owned automobile market is an ever-rising industry and have almost doubled its market value in the past decade. In today's world second hand cars have become very popular worldwide. The manufacturer sets the prices of new cars in the market, and the government imposes additional taxes. As a result, customers who purchase new cars can be confident that their investment is worthwhile. However, the high cost of new cars and customers' inability to afford them due to financial constraints have led to a rise in global sales of used cars (Arora et al., 2022). People belonging to middle class status who cannot afford to buy brand new, expensive cars can buy used cars nowadays. As a result of this, pre-owned car selling has increased to a large extent. Because of the proliferation of internet marketplaces like CarDheko, Quikr, Carwale, Cars24, and many more, it is now easier than ever for both the buyers and sellers to learn about the factors that affect a used car's price (Hankar et al., 2023). An efficient method is required to accurately evaluate the value of used cars by considering various features. While there are many websites that provide this service, their prediction techniques may not be ideal. Moreover, the effectiveness of predicting the actual market value of a used car may vary based on the model and system used. Therefore, it is very crucial to know the actual market value for used cars while purchasing or selling pre-owned cars. Generally the price of an used car is less than that of the original price of a car. Thereby, estimating the values of used cars is a very tedious job, as it depends on multiple factors like car mileage (number of kilometers travelled), manufacturing year, engine size, transmission type, power of the car and several other factors.

But, nowadays with the advent of modern technology like artificial intelligence, the retail value of an automobile can be estimated by applying various Machine Learning algorithms based on a predefined set of characteristics. There is no standard formula for estimating the selling price of used cars since different websites employ different algorithms to do so. One can easily get an approximate estimate of the price without actually entering the vehicle specifications into the desired website by training statistical models for forecasting the costs. The primary purpose of this research is to employ different ML prediction models to estimate the resale value of a used car and compare their performance accuracy parameters. Consequently, this results in substantial time and effort savings for both sellers and buyers interested in second-hand vehicles. Furthermore, the proposed model can also predict the variation in used car prices corresponding to different body types with respect to their manufacturing year. In addition, the car manufacturers such as Mercedes-Benz, Toyota, and Honda can determine which model should be produced in greater quantities if they wish to maintain competition in the used cars market.

2. Related Works

Pudaruth (2014) have proposed the prediction of the price of used cars by employing four different types of Machine Learning algorithms namely, Multiple Linear Regression Analysis, Naïve Bayes, Decision Trees and K-Nearest Neighbours. Pal et al. (2019) have proposed the methodology for car price prediction using Random Forest. In this paper, it has been concluded that good accuracy has been achieved from Random Forest in comparison to other previous works. Shanti et al. (2021) have proposed the idea of Machine Learning-Powered app for the prediction of prices of used cars. Four models were evaluated namely Random Forest, Neural Network, Gradient Boosting and Support Vector Regressor.

Venkatasubbu and Ganesh (2019) have estimated the used cars price prediction using Supervised learning techniques. In this paper using Lasso Regression, Regression trees and Multiple Regression, a statistical model was developed which based upon a given set of features and previous consumer data, the price of used cars were predicted. Amik et al. (2021) have estimated the application of machine learning techniques for prediction of cars which are pre-owned in Bangladesh. From this paper it has been concluded that XGBoost predicts the resale prices of used cars with higher accuracy.

AlShared (2021) have estimated the used cars price prediction and valuation using Data Mining techniques. This paper mainly predicts the price of used cars in Dubai. From this paper Random Forest has an accuracy of 95% which is the highest among all. Arefin (2021) have estimated Second Hand Price Prediction for Tesla Vehicles. This paper mainly stated that for the price prediction of a Tesla vehicle, how machine learning techniques such as SVM, Random Forest and deep learning techniques have been implemented.

Salim and Abu (2020) have developed a model namely, S-curve based on the used cars which have the maximum prices that are predictive in nature. To formulate maximum equation model of a new S-curve model, S-shaped Membership Function have been used as a base function. Farrell (1954) have discussed about the motor cars which have demand in the United States.

Monburinon et al. (2018) have predicted the used car prices by using Regression Models. Using supervised machine learning models, a relative study on regression performance had been conducted where Multiple Linear Regression, Random Forest Regression, Gradient Boosted Regression trees have been used to build used car's price model. By using Mean Absolute Error (MAE) as a parameter, the results were compared. Sun et al. (2017) have estimated the price evaluation model in Second-hand Car System based on the theory of BP Neural Network. A model of second-hand car price evaluation in online have been developed locally which helps in enhancing the speed and accuracy.

3. Proposed Methodology

In the current research problem, a prediction model is constructed by implementing various machine learning algorithms for predicting the prices of pre-owned cars by considering different parameters using regression analysis. The architecture of the proposed system is depicted in Figure 1 below.

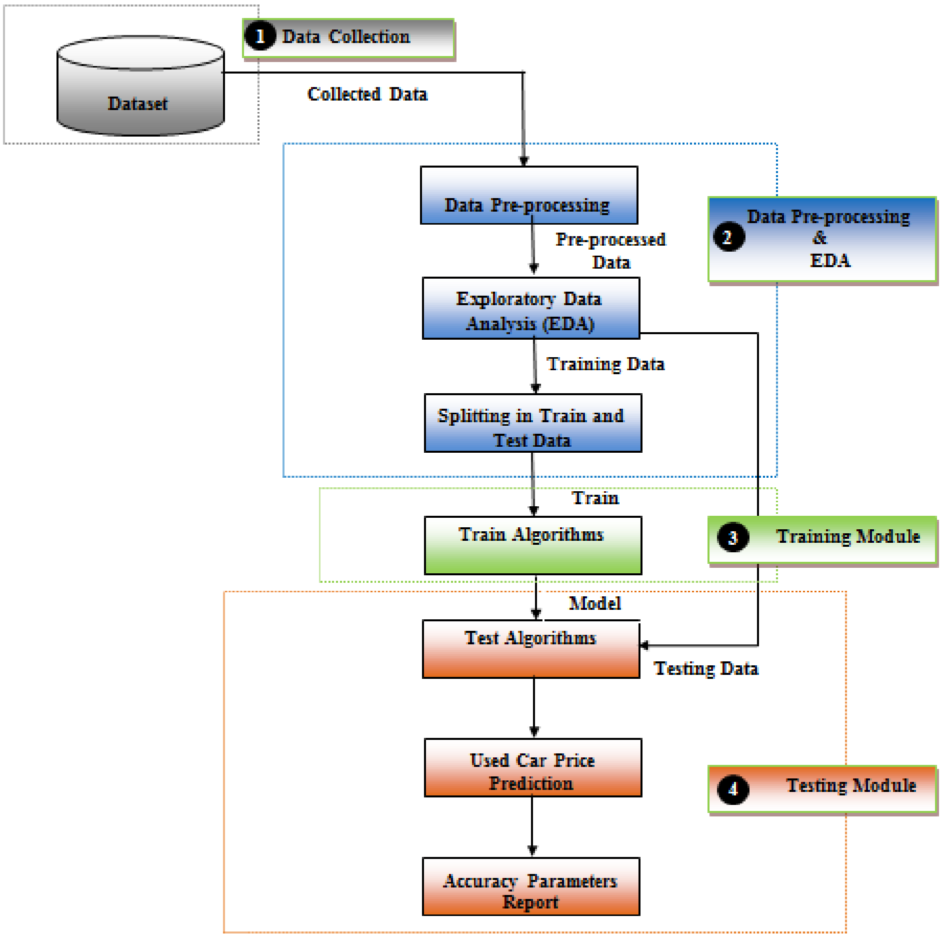

Figure 1. Control Flow Graph of the Proposed System

The proposed methodology is defined as follows:

- **Data acquisition**: At first data of different cars have been collected including features and target containing price as the main parameter.

- **Data cleaning**: Data cleaning comprises of identifying null values and removing them, filling missing values and removing outliers.
- **Preprocessing**: The preprocessing is being performed through Normalization or Standardization.
- **Exploratory Data Analysis (EDA):** Exploratory Data Analysis involves conducting initial investigations on data to identify patterns, detect anomalies, test hypotheses, and verify assumptions through the use of summary statistics and graphical representations.
- **Dividing into training and testing set**: The dataset which is obtained after preprocessing is being split into testing and training dataset.
- **Model training**: After the dataset is split into training and testing features the model is trained with the help of different machine learning algorithms by employing regression techniques.
- **Making predictions on the testing dataset**: The testing dataset which is obtained is being predicted, after that the testing values which are obtained is compared with the predicted values as a result of which price can be predicted.

In the next section, each of these points will be illustrated with respect to the results obtained.

4. Modeling and Result Analysis

4.1. Data Acquisition

The dataset employed in the current study has been downloaded from Kaggle and comprises of 426880 used cars scraped data records which is the world's largest collection of used vehicles for sale in the United States.

4.2 Data Cleaning

The data cleaning mainly comprises of removing the irrelevant features from the dataset namely, 'URL', 'region_url', 'vin', 'image_url' etc. to name a few of them. After that, the null values in the data are identified for each feature followed by the filling up of the missing values by applying appropriate methods and finally, removing the outliers from the data. Figure 2 displays the missing values before data cleaning process and Figure 3 depicts the corresponding null values in grey colour.

```
id                   0
region               0
year              1117
manufacturer     20747
model             6199
condition       186806
cylinders       166384
fuel              2991
odometer         75148
title_status      1806
transmission      2146
drive           122011
size            295961
type            117108
paint_color     135247
lat               8235
long              8235
price                0
```

Figure 2. Missing values before data cleaning process

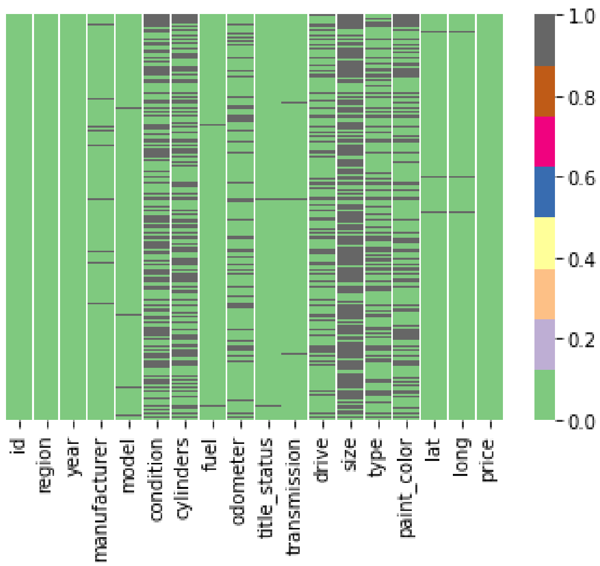

Figure 3. Distribution of null values in grey colour

To replenish the missing values in the data, the IterativeImputer technique is employed, with a variety of estimators being developed and their respective MSEs being generated using cross_val_score. Mean Squared Error (MSE) is computed as the average squared deviation between the true value and the predicted value retrieved from the data set. To deal with missing values, generally the MSE values are calculated by employing some central tendency measures like mean, median etc. along with some iterative imputation estimators. The imputation estimators employed in the current study are BayesianRidge Estimator, DecisionTreeRegressor Estimator, ExtraTreesRegressor Estimator and KNeighborsRegressor Estimator respectively. Figure 4 displays the MSE with 4 different Imputation methods.

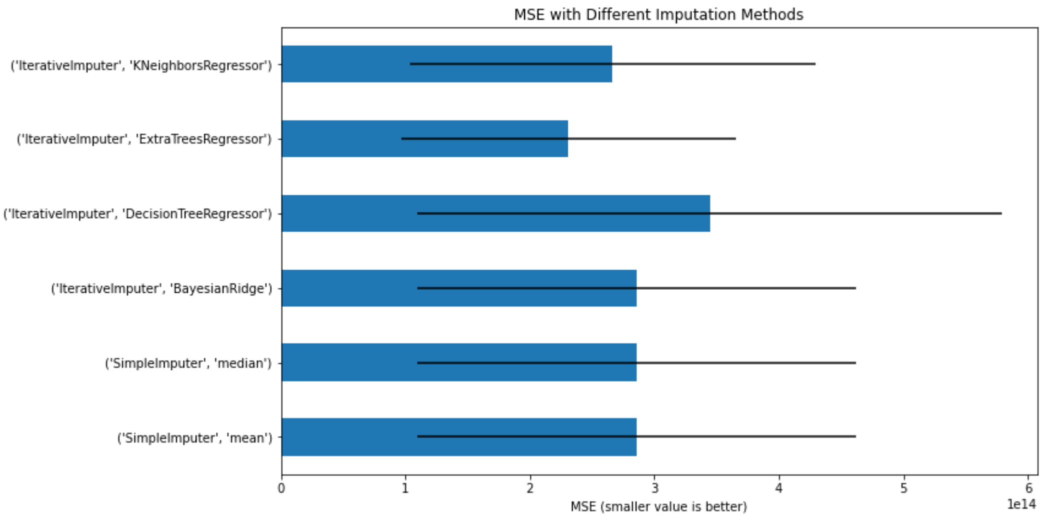

Figure 4. MSE with 4 different Imputation Methods

The preceding diagram suggests that the ExtraTreesRegressor estimator is preferable for the imputation strategy in the case of missing value. The second step is to fill up the missing values of categorical variables. One hot encoding is employed for this which converts each distinct value of a variable effectively into its corresponding binary variable. Figure 5 and 6 depicts the missing values after being filled up at the end of data cleaning process.

```
id              0
region          0
year            0
manufacturer    0
model           0
condition       0
cylinders       0
fuel            0
odometer        0
title_status    0
transmission    0
drive           0
size            0
type            0
paint_color     0
lat             0
long            0
price           0
```

Figure 5. Missing values after data cleaning process

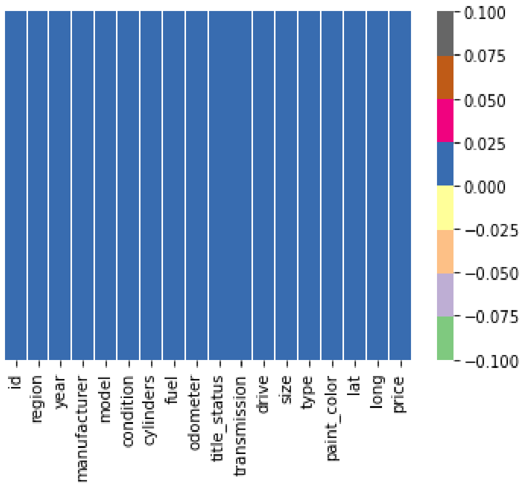

Figure 6. No null values in the dataset

The third and the final step is to remove the outliers from the data by employing the InterQuartileRange (IQR) method. Figure 7 and 8 depicts the Box Plots of Price and Odometer to reveal the outliers that lie within them. The price outliers in Figure 7 are those that have a logarithmic value less than 6.55 or greater than 11.55. Since no clear conclusion can be drawn from Figure 8, the interquartile range (IQR) is computed to identify the outliers, specifically for odometer values that fall below 6.55 or above 11.55.

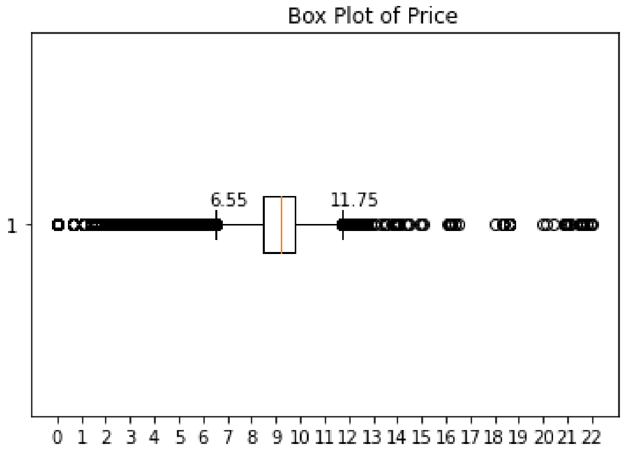

Figure 7. Box Plot of Price with outliers

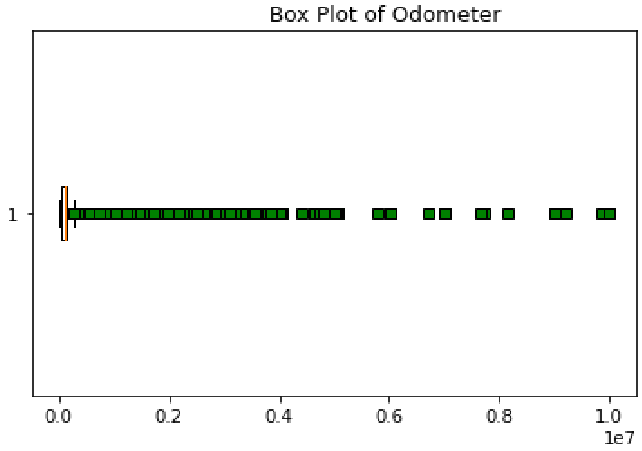

Figure 8. Box Plot of Odometer with outliers

Figure 9 displays the Box Plots and Histogram corresponding to Year. From Figure 9, it can be observed that the outliers are the year earlier than 1995 or later than 2020.

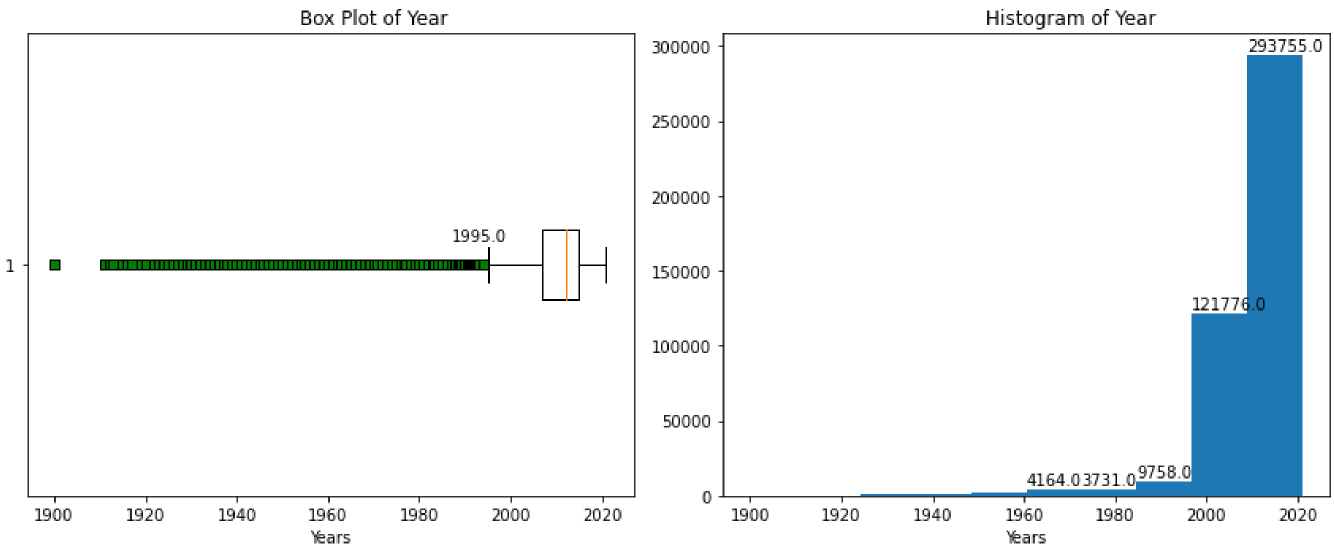

Figure 9. Box Plot and Histogram Plot of Year

Finally, after processing the dataset, its shape changed from (435849, 25) to (374136, 18), indicating that a total of 61713 rows and 7 columns were eliminated.

4.3. Data pre-processing

Data pre-processing is achieved through the implementation of Label Encoder and Normalization techniques.

- **Label Encoder:** The dataset comprises of 12 features which are categorical variables and 4 features which are numerical variables (excluding the price column). To utilize machine learning models, it is mandatory to convert these categorical variables into numerical variables. The sklearn library's LabelEncoder is being utilized to accomplish this task.
- **Normalization:** The dataset is not distributed normally and each feature has a distinct range. If the data is not normalized, the machine learning model may ignore features with low values as their impact will be negligible compared to the larger values. To overcome this issue, the sklearn library's MinMaxScaler is utilized to normalize the data.

4.4 Exploratory Data Analysis

Let us now explore the various Exploratory Data Analysis (EDA) visualizations in the current dataset. Figure 10 depicts the correlation plot among the various feature variables in the dataset.

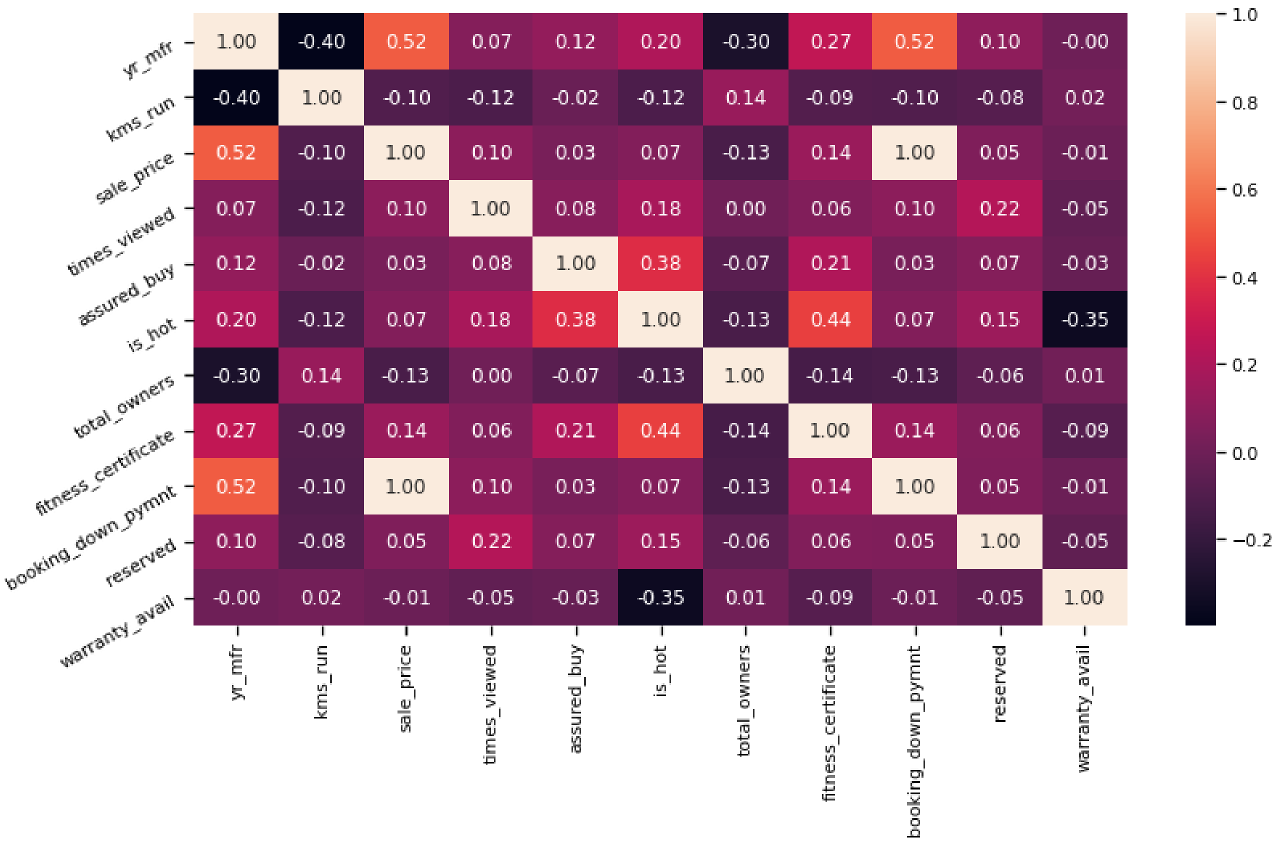

Figure 10. Correlation Matrix Plots among the different variables

It can be noted that there is low correlation among the features present in the data. Next, the pair-plots between the various variables is illustrated in Figure 11.

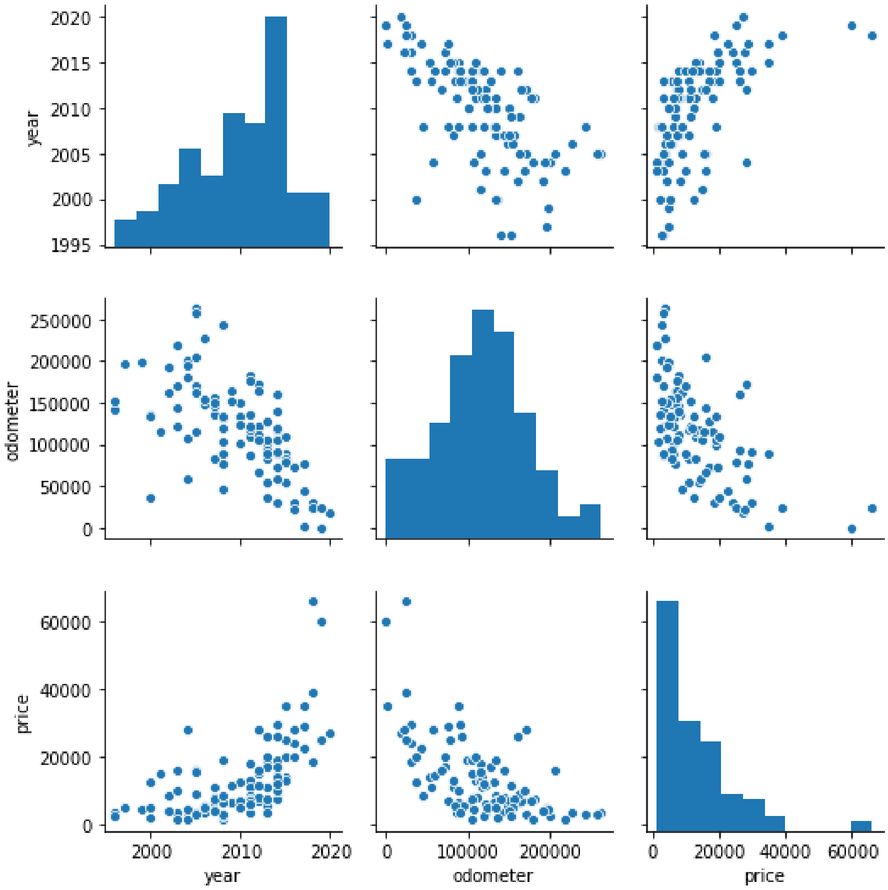

Figure 11. Pair-plots to find correlation

The pair plot doesn't provide any conclusive evidence as there is no apparent correlation between the variables. Figure 12 represents the distribution of price.

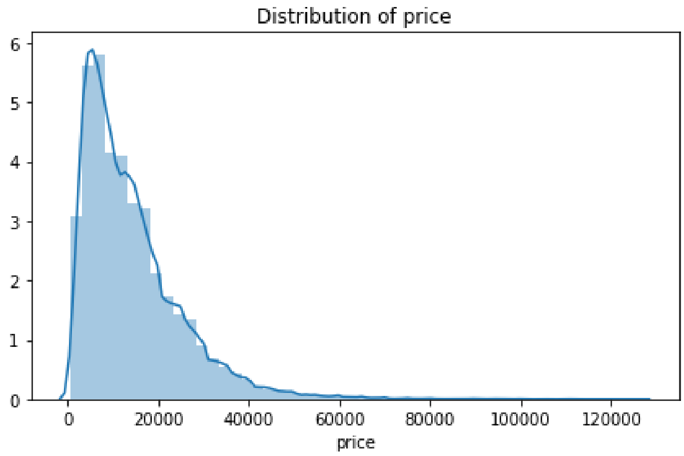

Figure 12. Graph showing distribution of Price

Based on the information as displayed in the Distplot, it can be inferred that the price undergoes a rapid hike in the beginning, but after a certain time, it begins to depreciate. Next, Figure 13 describes the bar plot of price plots corresponding to each fuel type.

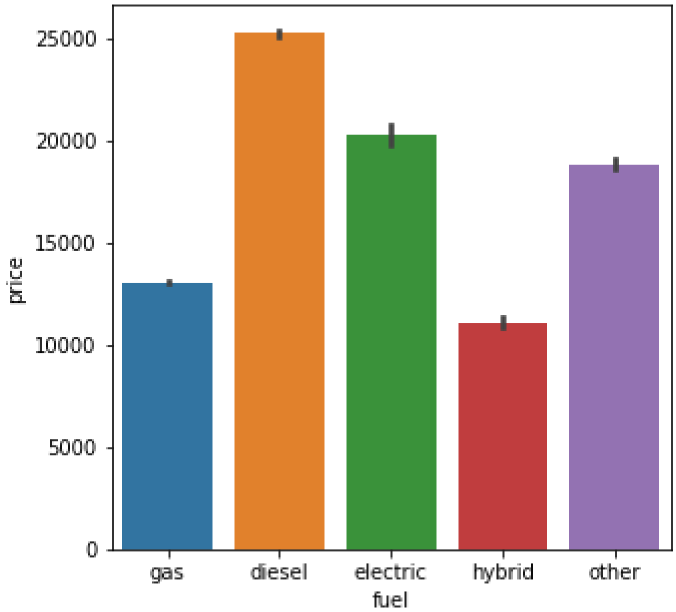

Figure 13. Bar Plots displaying the price of each fuel type

Upon analysis of the graph, it can be concluded that the cost of diesel cars is higher than that of electric cars, while hybrid vehicles are the least expensive. Figure 14 depicts the variation of car price and fuel type with change in hue condition.

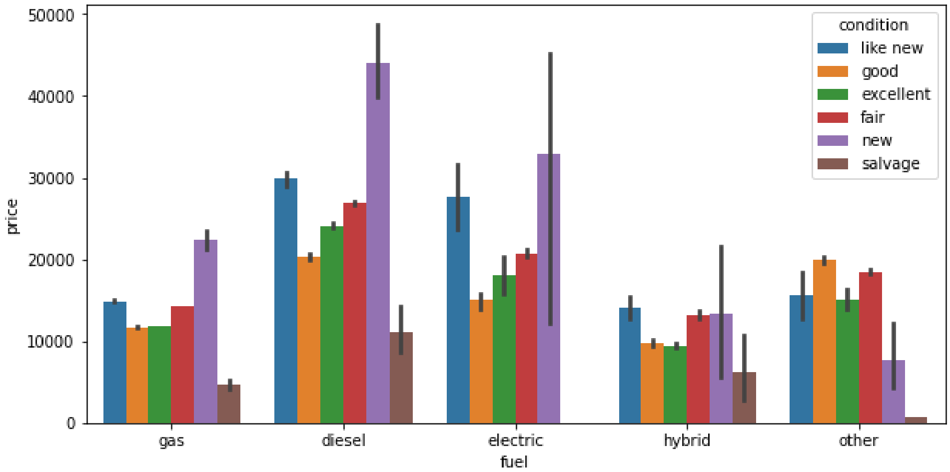

Figure 14. Bar Plots of fuel and price with hue condition

From this bar-plot analysis, it can be concluded that the hue condition of a car also plays a significant role in determining its price based on the type of fuel it uses. Figure 15 depicts the car prices variation with year.

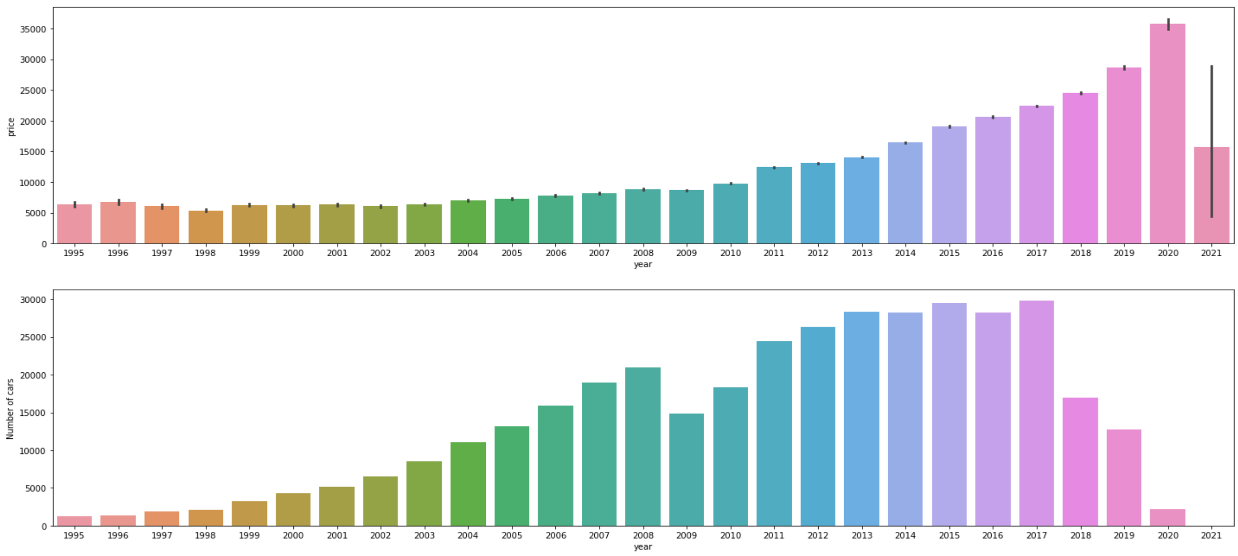

Figure 15. Graph displaying car price variation per year

The first plot in Figure 16 indicates that the prices of cars have been consistently rising annually since 1995, while the second plot illustrates an increasing trend in the number of cars per year. However, it can be observed that there is a point in time, specifically in 2012, where the number of cars seems to plateau and remain relatively constant.

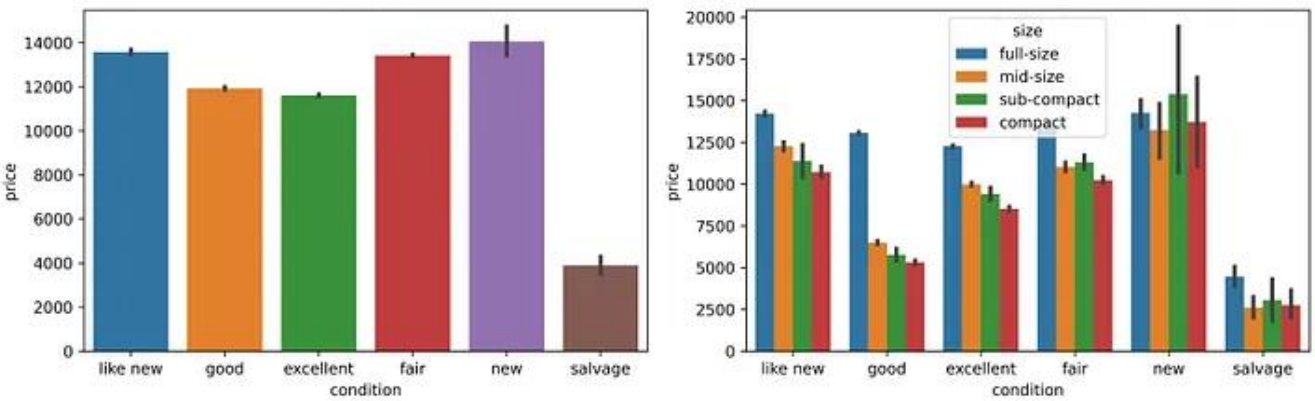

Figure 16. Bar Plot displaying the price with respect to the car condition

From Figure 16, it can be deduced that the price of cars is influenced by their condition, as the car price fluctuates according to the car's size and condition. Figure 17 depicts the car price with respect to transmission type.

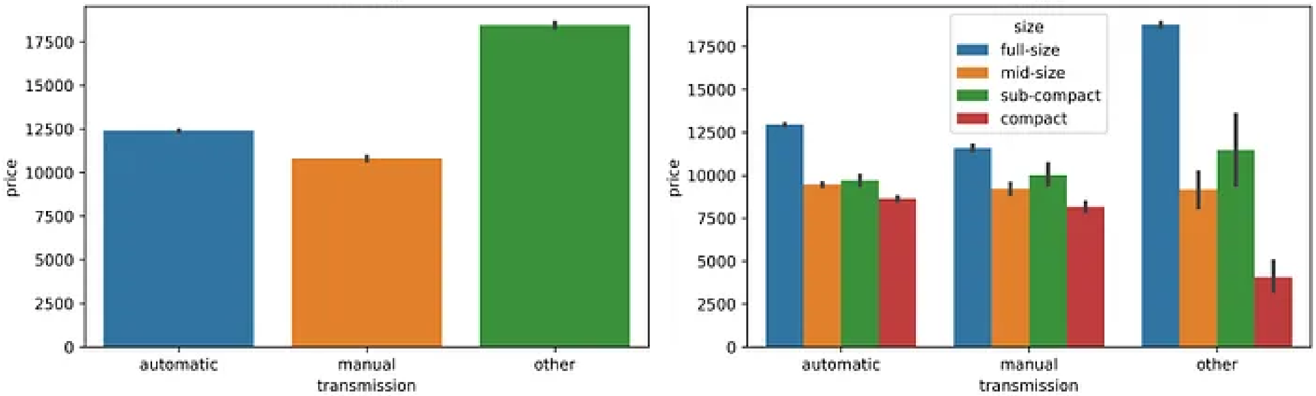

Figure 17. Bar Plot displaying the price with respect to the car transmission type

Upon analysis, it is evident that the price of cars differs depending on the type of transmission. Buyers are willing to purchase cars with automatic transmission, while cars with manual transmission are priced lower.

4.5 Splitting the dataset into training and testing set

During this procedure, 90% of the data was allocated for the training dataset, while the remaining 10% was designated as the testing dataset.

4.6 Training with ML Models

This section involves utilizing various machine learning algorithms to predict the the price of pre-owned cars which is the target variable in the current study. We now apply a set of supervised machine learning algorithms to achieve the targeted pre-owned car prices.

(a) Linear Regression: Linear regression is a statistical technique that involves modeling the correlation between a single outcome variable, also known as the dependent variable, and one or more explanatory variables, also known as independent variables. Linear predictor functions are utilized to model the relationships in linear regression, and their unknown parameters are estimated from the data. These models are referred to as linear models. The coefficients in a statistical model show whether the relationship between a predictor variable and the response variable is positive or negative.

- A positive coefficient implies that an increase in the predictor variable leads to an increase in the response variable.
- A negative coefficient implies that an increase in the predictor variable leads to a decrease in the response variable.

Figure 18 illustrates the performance of Linear Regression algorithm and Figure 19 displays the predominant features using the Linear Regression algorithm.

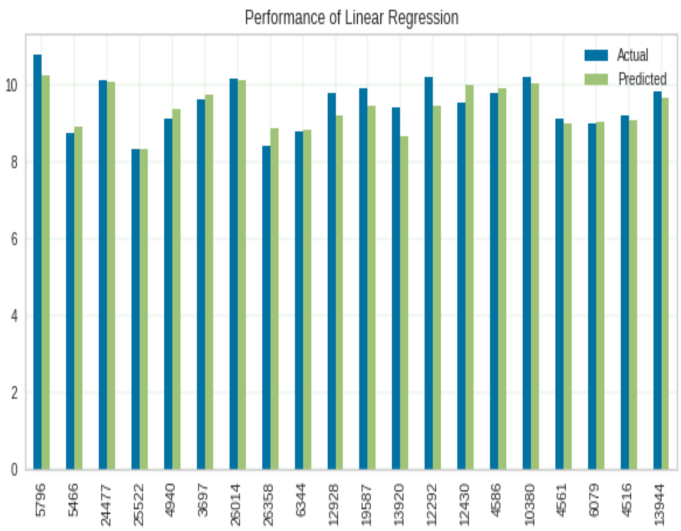

Figure 18. Graph displaying performance of LR

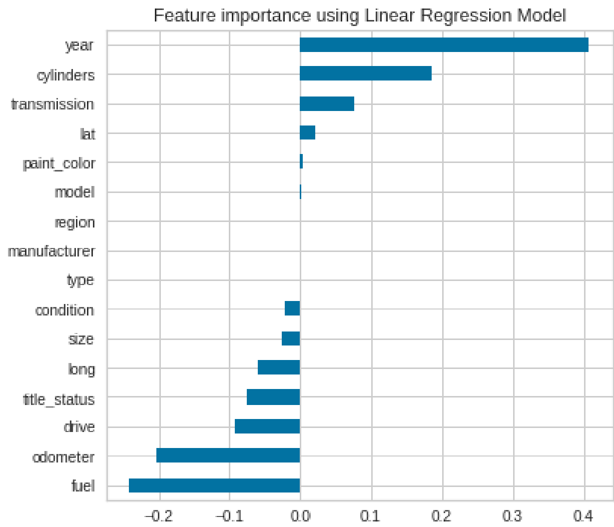

Figure 19. Feature importance using LR

Based on the graph, it can be inferred from the Linear Regression analysis that the year, cylinder, transmission, fuel, and odometer variables are the most significant ones.

(b) Ridge Regression: Ridge Regression is a method used to examine multiple regression data that is affected by multicollinearity. In cases of multicollinearity, the least squares estimates may be unbiased, but their variances are substantial, which can result in values that are significantly different from the actual ones.

In order to determine the optimal alpha value for Ridge Regression, the AlphaSelection tool from the yellowbrick library was utilized.

Figure 20 displays the Ridge Regression alpha error while Figure 21 represents the feature importance corresponding to Ridge model.

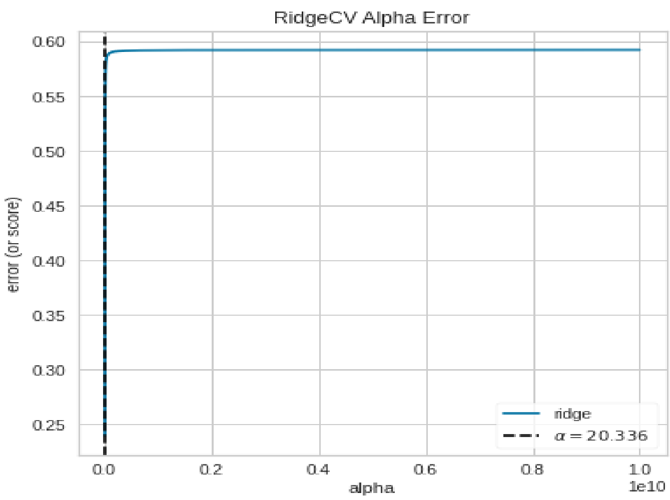

Figure 20. Graph displaying best value of Alpha

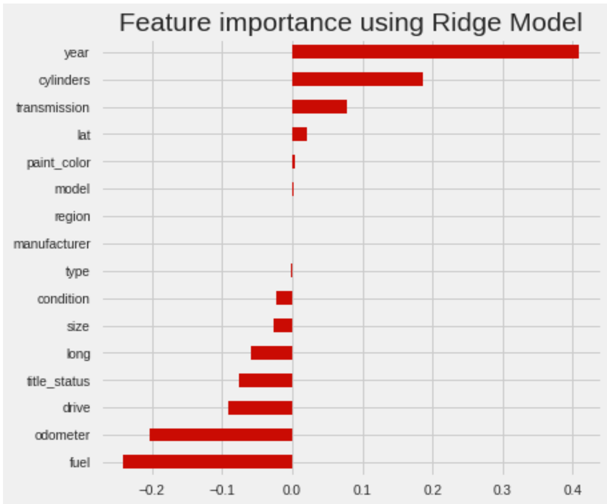

Figure 21. Feature importance using RR

According to the figure plotted, the optimal alpha value for adjusting the dataset is 20.336. It should be noted that alpha value is not fixed and can change each time. The Ridge Regressor method is applied based upon this alpha value. The figure also suggests that year, cylinder, transmission, fuel and odometer are the most prominent feature variables.

(c) Lasso Regression: Lasso Regression is a form of linear regression that implements shrinkage, which involves pulling data values towards a central point such as the mean. By using the Lasso approach, the development of straightforward, concise models is promoted. The objective of Lasso Regression is to identify the subset of predictors which results in the lowest prediction error for a quantitative response variable. To achieve this, the Lasso applies a restriction on the model parameters that induces regression coefficients for certain variables to contract to zero value.

Figure 22 depicts the most prominent features corresponding to Lasso model.

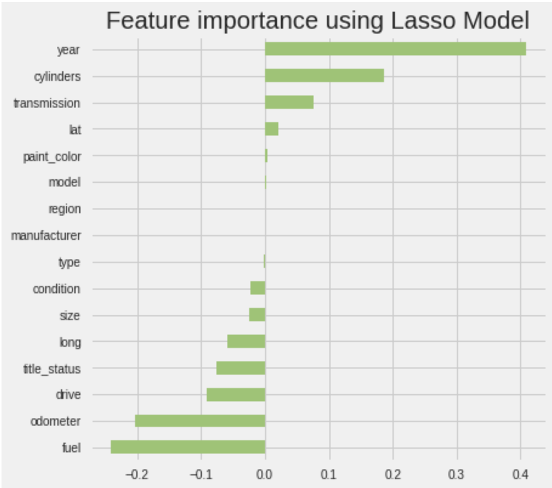

Figure 22. Lasso Regression feature importance

(d) K-Nearest Neighbor: Local interpolation is used to predict the target based on the nearby targets in the training set. This approach is known as K-NN, which is a form of lazy learning or instance-based learning. In k-NN, the function is only approximated locally, and all calculations are postponed until function evaluation.

Figure 23 represents the error plot of k-NN model. The figure depicts the least error for k=5 with n_neighbors=5 and metric='euclidean'. Thus, it can be concluded that the performance of KNN is better since as the accuracy increases, the error decreases.

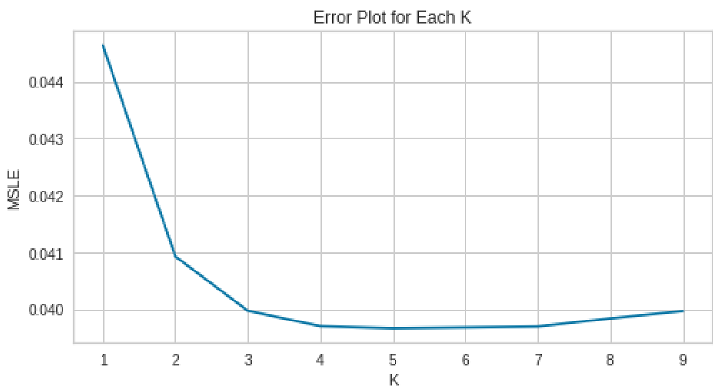

Figure 23. Error Plot for KNN for k range 1-9

(e) Random Forest: The Random Forest is a classification technique that involves multiple decision trees. To generate a diverse set of trees with uncorrelated predictions, the algorithm employs bagging and feature randomness during tree construction. By aggregating the predictions of all the trees, the Random Forest algorithm aims to produce more accurate results than any single decision tree. Our model generates 180 decisions by implementing a maximum of 50% of the available features.

Figure 24 and Figure 25 describes the performance and feature importance of Random Forest classifier respectively.

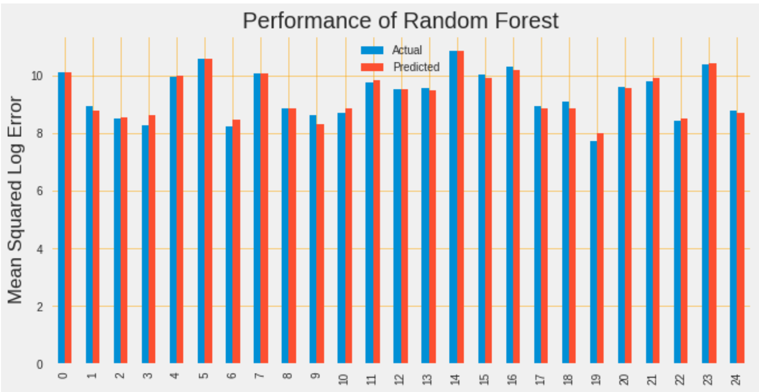

Figure 24. Performance Graph of RF

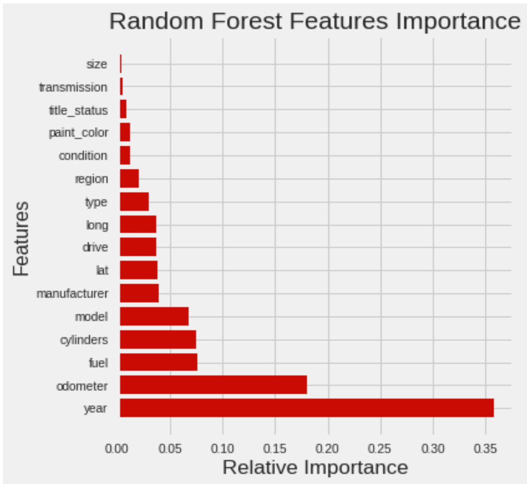

Figure 25. Feature importance using RF

The basic bar chart demonstrates that the year of the car is the most significant characteristic, followed by the odometer variable and then other variables. The Random Forest algorithm has displayed improved performance with an increase in accuracy of approximately 10%, which is positive. As the algorithm utilizes bagging in building each tree, so the next step will be to perform the Bagging Regressor.

(f) Bagging Regressor: A Bagging Regressor is a type of ensemble meta-estimator that builds individual regression models on random subsets of the original dataset and then combines their predictions to produce a final prediction. This can be executed by taking a vote or by averaging the individual predictions. The purpose of this meta-estimator is to decrease the variability of a black-box estimator, such as a decision tree, by adding randomness to its creation process and then creating an ensemble from it.

(g) AdaBoost Regressor: AdaBoost is a machine learning technique that can enhance the effectiveness of any other machine learning algorithm. By combining several "weak classifiers" into a single "strong classifier," AdaBoost assists in this process. Figure 26 describes the feature importance of AdaBoost classifier. A quick look at the bar chart reveals that year is the most influential factor, followed by the total mileage driven, and then model etc.

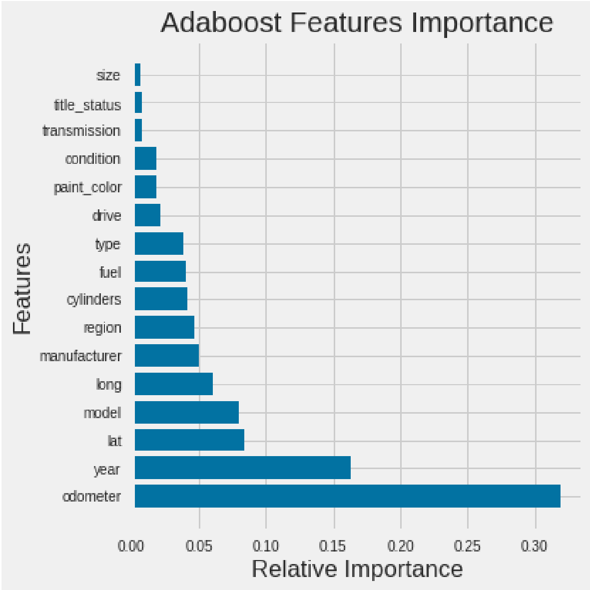

Figure 26. Feature importance of AdaBoost classifier

(h) XGBoost Regressor: XGBoost is a method of ensemble learning that utilizes gradient boosted decision trees. Its key advantage is its ability to quickly and efficiently learn through parallel and distributed computing, as well as its effective use of memory. This powerful algorithm's scalability is what makes it such an attractive option for many applications. Figure 27 illustrates the feature importance of XGBoost classifier.

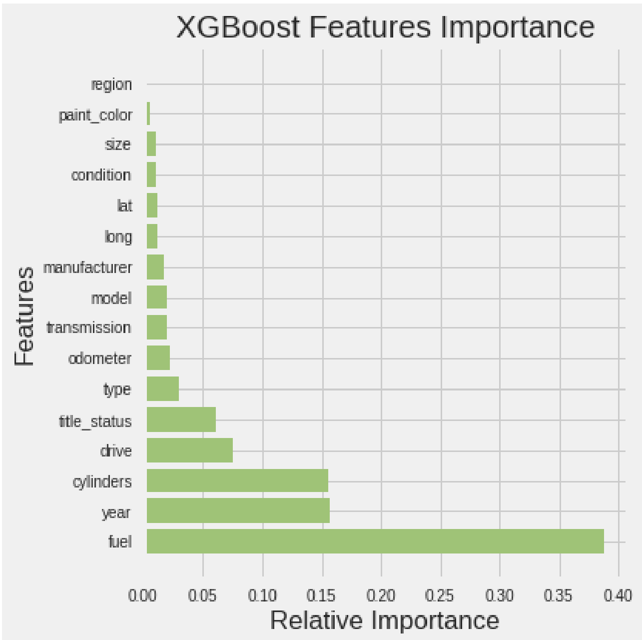

Figure 27. Feature importance of XGBoost classifier

The bar plot is a straightforward representation that ranks the car features in order of their importance, showing which ones carry more weight. XGBoost analysis indicates that the Odometer is a significant feature, while in earlier models, the year was identified as an essential factor.

4.7 Comparison of the Performance of ML Models

Assessing the accuracy of a machine learning model is a crucial step in developing it, as it helps determine how effective the model is in making predictions. The most commonly used metrics for evaluating the model's performance and prediction error rates in regression analysis are MSLE, MSE, MAE, RMSE, and R2 Score which are defined as follows:

- **MSLE**: The loss function referred to as MSLE (Mean Squared Logarithmic Error) is designed to alleviate the harsh impact of large discrepancies in high predicted values. This makes it a more suitable choice for evaluating models that make predictions directly in unscaled quantities.
- **RMSLE**: Also, known as Root Mean Squared Logarithmic Error, the RMSLE is calculated by taking the square root of the mean of the squared differences between the natural logarithms of the predicted and actual values that have been transformed to a logarithmic scale. To prevent the calculation of the natural logarithm of values that could be zero, 1 is added to both the actual and predicted values before the transformation is performed.
- **MAE**: Mean absolute error (MAE) is a metric that calculates the average absolute difference between the predicted values and the actual values in a dataset. It is obtained by taking the absolute difference between each predicted value and its corresponding actual value, averaging these differences, and then expressing the result as a single value.
- **MSE**: Mean squared error (MSE) is a measure of the average squared difference between the predicted values and the actual values in a dataset. It is computed by taking the squared difference between each predicted value and its corresponding actual value, averaging these squared differences, and then expressing the result as a single value.
- **RMSE**: Root mean squared error (RMSE) is a metric that represents the square root of the MSE (mean squared error). It is calculated by taking the square root of the average squared difference between the predicted values and the actual values in a dataset.
- **R2 Score**: The R2 score, also known as the coefficient of determination, is utilized to assess the effectiveness of a linear regression model. Its purpose is to determine how accurately the model replicates observed outcomes by calculating the proportion of the overall variation in results that can be explained by the model.

Now, let us have a close insight regarding the performance of the various machine learning models implemented on the used car dataset. Figure 28 represents the accuracy parameters of the various ML models implemented in the current study and Figure 29 depicts the diagrammatic representation of the highest accuracy achieved by different ML models.

	Linear Regression	Ridge Regression	Lasso Regression	KNN	RandomForest Regressor	Bagging Regressor	AdaBoost Regressor	XGBoost Regressor
MSLE	0.002427	0.002427	0.002427	0.001573	0.000825	0.001525	0.000832	0.000818
Root MSLE	0.049264	0.049264	0.049264	0.039666	0.028718	0.039056	0.028846	0.028607
MAE	0.363932	0.363934	0.363938	0.264662	0.167132	0.236025	0.173273	0.162709
MSE	0.242811	0.242811	0.242808	0.154488	0.078840	0.148183	0.080097	0.078217
RMSE	0.242811	0.242811	0.242808	0.154488	0.078840	0.148183	0.080097	0.078217
R2 Score	0.592516	0.592516	0.592520	0.740738	0.867691	0.751319	0.865582	0.868736
Accuracy(%)	59.251563	59.251570	59.252033	74.073797	86.769137	75.131900	86.558152	86.873640

Figure 28. Accuracy Parameters of ML models

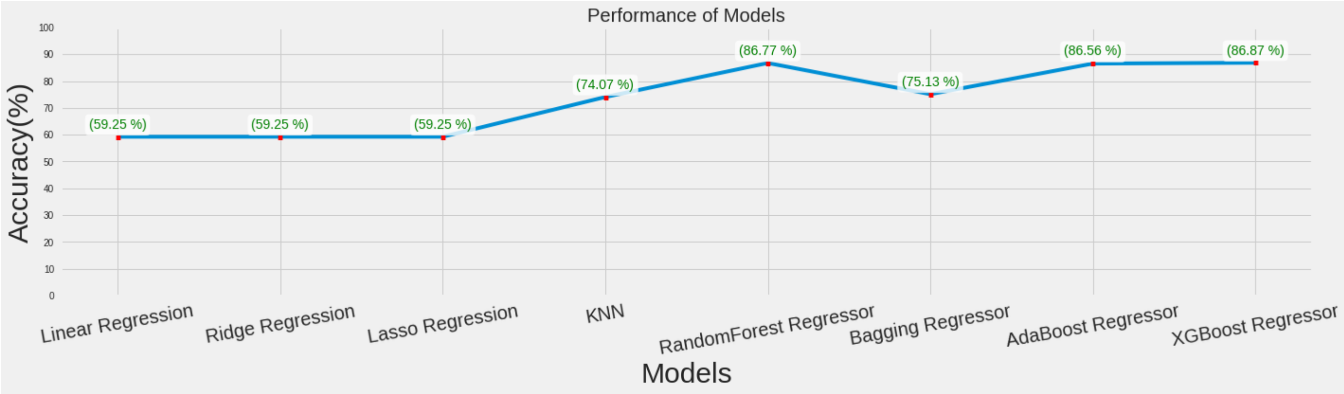

Figure 29. Performance of various ML models

Based on the figure presented above, we can infer that the XGBoost regressor has a higher level of performance than the other models, with an accuracy of 86.87%.

5. Conclusion

The objective is to predict the price of used cars by employing 25 predictors. To achieve the highest possible accuracy and minimize errors, various machine learning models were evaluated.

At first, the dataset underwent data cleaning to eliminate any null values or outliers. Subsequently, machine learning models are employed to make predictions about car prices. Then, using data visualization tools, a thorough examination of the features is conducted to investigate the relationships between them. Based on the table provided, it can be inferred that XGBoost is the most suitable model for forecasting used car prices. XGBoost, employed as a regression model, demonstrated the most optimal MSLE and RMSE outcomes.

This work also proposes a future scope where deep learning algorithms on the same dataset to get more accurate results with higher efficiency. Also, other datasets can be utilized for a comparative study.

References

AlShared, A. (2021). Used Cars Price Prediction and Valuation using Data Mining Techniques. Thesis. Rochester: Rochester Institute of Technology.

Amik, F. R., Lanard, A., Ismat, A., & Momen, S. (2021). Application of Machine Learning Techniques to Predict the Price of Pre-Owned Cars in Bangladesh. Information, 12(12), 514.

Arefin, S. E. (2021). Second Hand Price Prediction for Tesla Vehicles. arXiv:2101.03788

Arora, P., Gupta, H., & Singh, A. (2022). Forecasting resale value of the car: Evaluating the proficiency under the impact of machine learning model. Materials Today: Proceedings, 69 (2), 441-445.

Farrell, M. J. (1954). The demand for motor-cars in the United States. Journal of the Royal Statistical Society. Series A (General), 117(2), 171-201.

Hankar, M., Birjali, M., & Beni-Hssane, A. (2023). Machine Learning Modeling to Estimate Used Car Prices. In Innovations in Smart Cities Applications Volume 6: The Proceedings of the 7th International Conference on Smart City Applications (pp. 533-542). Springer.

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

Pal, N., Arora, P., Kohli, P., Sundararaman, D., Palakurthy, S. S. (2019). How Much Is My Car Worth? A Methodology for Predicting Used Cars' Prices Using Random Forest. In: Arai, K., Kapoor, S., Bhatia, R. (eds)

Advances in Information and Communication Networks. FICC 2018. Advances in Intelligent Systems and Computing, vol 886 (pp. 413-422). Cham: Springer, Cham.

Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. International Journal of Information and Computer Technology, 4(7), 753-764.

Salim, F., & Abu, N. A. (2020). An S-curve model on the maximum predictive pricing of used cars. European Journal of Molecular and Clinical Medicine, 7(3), 907-921.

Shanti, N., Assi, A., Shakhshir, H., & Salman, A. (2021). Machine Learning-Powered Mobile App for Predicting Used Car Prices. The Proceedings of the 3rd International Conference on Big-data Service and Intelligent Computation (BDSIC 21) (pp. 52-60). New York: Association for Computing Machinery.

Sun, N., Bai, H., Geng, Y., & Shi, H. (2017). Price evaluation model in second-hand car system based on BP neural network theory. The Proceedings of the 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 431-436). Kanazawa: IEEE.

Venkatasubbu, P., & Ganesh, M. (2019). Used Cars Price Prediction using Supervised Learning Techniques. International Journal of Engineering and Advanced Technology, 9(1S3), 216-223.