Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Predicting Fair Pricing in the Used Car Market: A Machine Learning Approach Using PyCaret

Ashish Kumar Soni[1], Pooja Patre[1], Jasmine Minj[1], Divya Shukla[2], and Advait Khare[3]

[1]Department of Computer Science and Engineering, Vishwavidyalaya Engineering College Ambikapur, India
[2]Department of Computer Science and Engineering, Chhattisgarh Swami Vivekanand Technical University, Bhilai, India
[3]Department of Computer Science and Engineering, Shri Shankaracharya Institute Of Professional Management and Technology, Raipur, India
Email: ashishsoni088@gmail.com

In the current era of renewable energy vehicles, such as electric cars, government support has significantly influenced the used car market, making it essential to analyze trends in car pricing. Determining whether the price of a used car is fair remains a crit- ical challenge for buyers and sellers, especially given the variety of car specifications and market fluctuations. To address this, machine learning methods can be utilized to predict car prices effectively and assist decision-making.

This study leverages a dataset comprising 205 records, featuring car attributes such as fuel type (e.g., gas, diesel), car body type (e.g., sedan, convertible, hatchback), en- gine specifications (e.g., size, type, and number of cylinders), and performance indi- cators (e.g., horsepower, mileage). The dataset includes a diverse range of cars with prices ranging from economy to luxury segments, enabling a comprehensive analysis. Using PyCaret, an automated machine learning library, we compared the perfor- mance of two predictive models: Random Forest and Decision Tree. The methodology included data preprocessing, feature selection, and hyperparameter tuning for each model. The evaluation criteria focused on model accuracy, interpretability, and prac- tical applicability. The models exhibited distinct levels of accuracy, with the Random

Forest method outperforming the Decision Tree in overall predictive performance, achieving an accuracy of TFL%.

The findings highlight the potential of machine learning models in providing reliable car price predictions, emphasizing their role in enabling fair pricing in the used car market. Future work will explore the integration of additional factors such as market trends and economic indicators to further enhance prediction accuracy and usability.

1. Introduction

The used car market has evolved into a critical component of the global automotive in- dustry, serving as an accessible entry point for many consumers and a significant profit generator for businesses. The complexity of pricing used vehicles stems from the wide ar- ray of factors that influence their resale value. Variables such as the vehicle's make, model, year, mileage, and condition interact with market trends, regional demand, and economic conditions, creating a multifaceted problem for pricing predictions.

The prices of new cars in the market are typically determined by manufacturers, with additional government-imposed costs, such as taxes. This pricing structure assures cus- tomers of the value of their investment when purchasing a new car. However, the rising costs of new vehicles and the financial constraints faced by many buyers have led to a sig- nificant increase in global demand for used cars. Predicting the prices of used cars has become an important and challenging problem that requires attention. Many customers are often exploited through unrealistic pricing of used cars, emphasizing the need for a reliable system to accurately determine their value based on a range of features.

In developed countries, the high prices of cars and the mobility-driven lifestyles of individuals have popularized leasing agreements, where cars are resold after the lease pe- riod ends. As a result, the resale of cars has become an integral part of the automotive market. Predicting the price of a used car is not straightforward due to the multitude of factors influencing its value. Key factors include the car's age, make, country of origin, mileage, and horsepower. Rising fuel costs make fuel economy a critical consideration, along with other features such as fuel type, body style, braking system, engine size (mea- sured in cc), acceleration, and safety ratings. Additional attributes, including the number of doors, dimensions (size, weight, and height), paint color, consumer reviews, awards won by the manufacturer, and optional features such as sound systems, air conditioning, GPS navigation, and premium wheels, also play a role in determining a car's value.

This study proposes leveraging PyCaret, an open-source, low-code machine learning library, to explore a range of models for predicting the prices of used cars based on these attributes. The methodology involves preprocessing the dataset to ensure data quality, fol- lowed by utilizing PyCaret's automated machine learning pipeline to experiment with mul- tiple algorithms, including regression-based models. PyCaret simplifies model comparison and tuning, enabling the selection of the best-performing approach for price prediction.

The structure of this paper is as follows: Section 2 reviews related work on used car price prediction. Section 3 discusses the dataset and preprocessing steps. Section 4 presents the results of exploratory data analysis. Section 5 details the methodology, focusing on the use of PyCaret to evaluate multiple machine learning models and their performance. Finally, Section 6 concludes the study and outlines future work.

2. Related work

The study utilizes a dataset from Kaggle to predict used car prices. The dataset includes various features, as outlined in Section 3 of this paper, which are essential for predicting and classifying the price ranges of used cars. A review of related literature highlights several studies where researchers have used similar or related datasets for price prediction tasks.

In Pal et al., 2019, a patented engine platform for asset price assessment is described. This platform uses a price computation matrix and may apply a linear regression model to determine asset prices based on a set of input variables. However, the paper lacks specific details about features applicable to particular vehicle types. In contrast, our study identifies and incorporates critical features into a Random Forest model for predicting used car prices. Zhang et al. Zhang et al., 2017 used a Kaggle dataset to predict used car prices. The study evaluated the performance of multiple classification models, including logistic re- gression, SVM, decision trees, Extra Trees, AdaBoost, and Random Forest. Among these, the Random Forest classifier achieved the best performance, with an accuracy of 83.08% on the test data. Their model used five features (brand, powerPS, kilometer, sellingTime, and VehicleAge) after removing irrelevant features and outliers. Our approach builds on this by incorporating additional relevant features, such as car price and vehicle type, which signif- icantly impact price predictions. Furthermore, we broaden the range of features, including year of registration and powerPS, to enhance the model’s predictive accuracy.

Awad et al. Awad and ELseuofi, 2011 presented an educational paper reviewing six popular classification methods (Bayesian classification, ANNs, SVMs, k-NN, Rough Sets, and Artificial Immune Systems) for spam email classification. This work provided valuable insights into classification models but focused on a binary classification task. In contrast, our study employs a one-vs-rest approach, leveraging the Random Forest model to address feature dependency issues in our dataset. The hybrid approach suggested by Awad et al. aligns with our methodology, leading to improved accuracy.

Durgesh et al. Durgesh and Lekha, 2010 offered an introductory study on Support Vector Machines (SVMs), comparing their performance against other techniques like k-NN and rule-based classifiers using datasets from the UCI Machine Learning Repository. Their findings demonstrated that SVMs achieve superior classification accuracy, establishing a baseline for prediction tasks. However, our study progresses beyond simple linear models by utilizing Random Forest, which provides higher accuracy for used car price prediction. In Pudaruth, 2014, the authors predicted used car prices in Mauritius using four machine learning algorithms: multiple linear regression, k-nearest neighbors, naive Bayes, and de- cision trees. Their dataset was derived from historical data in local newspapers. Although their approach offered comparable results, it lacked the predictive accuracy achieved in our study using Kaggle data. By employing a sophisticated algorithm like Random Forest and preprocessing our dataset to include relevant features, our study demonstrates superior results.

Noor et al. Noor and Jan, 2017 applied multiple regression models to predict car prices using a dataset from PakWheels.com. Their dataset included approximately 2,000 records collected over a short duration and featured variables such as color and advertisement date, which were deemed less relevant for price prediction. In contrast, our study emphasizes se- lecting more impactful features, such as brand and mileage, resulting in improved accuracy using Random

Forest on the Kaggle dataset.

Finally, researchers in Kuiper, 2008 used a multivariate regression model to predict used car prices, employing a dataset of 2005 General Motors vehicles. Their study em- phasized variable selection techniques for identifying relevant features, providing valuable insights for similar tasks. However, their model did not require extensive preprocessing, relying on basic portal data (e.g., PakWheels.com). Our approach, by contrast, focuses on preprocessing Kaggle data, selecting relevant features, and generalizing the model for various brands and years, leading to more robust predictions.

3. Methodology

3.1 Data Description

This dataset contains information about cars collected from the American automobile mar- ket to help a Chinese automobile company, Geely Auto, understand the factors influencing car pricing in the United States. The dataset consists of 205 entries and 26 columns, with each row representing a car and each column representing a specific attribute of that car. The dataset is intended for analyzing and predicting car prices based on the above features. It aims to identify significant factors influencing car prices in the U.S. market and evaluate their impact to assist Geely Auto in strategizing their entry into the market.

Table 1: Key features in the dataset.

General Information	Insurance Risk	Physical Dimen- sions and Weight	Engine Spec- ifications	Performance and Economy	Pricing
car ID − CarName	symboling fueltype aspiration doornumber carbody drivewheel enginelocation	wheelbase carlength carwidth carheight curbweight	enginetype cylindernumber enginesize fuelsystem boreratio stroke compressionratio horsepower peakrpm	citympg highwaympg	price (target)

Table 1 lists the various features present in the dataset. The dataset captures a compre- hensive range of attributes essential for analyzing car prices in the American market. Each car is uniquely identified by the car ID and described by its CarName, which includes its make and model. Insurance risk is represented by the symboling score, which ranges from

-3 to 3, indicating increasing levels of risk with higher values. The dataset also provides general specifications, including the fueltype (e.g., gas or diesel), aspiration system (e.g., standard or turbo), doornumber (two or four), and carbody type (e.g., sedan, hatchback, or convertible). Additional details include the car’s drivewheel configuration (e.g., front- wheel drive, rear-wheel drive, or 4WD) and enginelocation (front or rear).

Physical dimensions and weight are captured through attributes such as wheelbase (dis- tance between the front and rear wheels), carlength, carwidth, carheight (all measured in inches), and curbweight (the car's weight without passengers or cargo). Engine specifications are described with features like enginetype (e.g., dohc, ohcv, or l), cylindernumber, enginesize (in cubic inches), and fuelsystem (e.g., mpfi, 2bbl). Additional technical details include the engine's boreratio (cylinder bore diameter), stroke (cylinder stroke length), compressionratio, horsepower, and peakrpm (maximum engine revolutions per minute).

Performance and fuel economy are captured through citympg (miles per gallon in city driving) and highwaympg (miles per gallon on highways). The target variable, price, repre- sents the market price of each car, serving as the key metric to be predicted in this analysis. These diverse attributes enable a detailed exploration of factors influencing car pricing in the U.S. market.

3.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial step in understanding the structure and underlying patterns within a dataset. For the given dataset on car prices, EDA involves examining the data distribution, identifying significant variables, detecting missing values or anomalies, and uncovering relationships between features. This analysis helps in determin- ing which factors most strongly influence car pricing, paving the way for effective feature engineering and model building.

Figure 1 depicts the density and frequency of car prices in the dataset. The plot reveals a positively skewed distribution, indicating that most cars are priced in the lower range, with a significant number of vehicles concentrated around the $10,000 mark. The distribution tapers off as prices increase, with fewer cars in the higher price range beyond $30,000. This suggests a market dominated by affordably priced vehicles, while luxury or high-end models form a smaller proportion of the data. Understanding this distribution is critical for designing models that can predict prices accurately across different segments.

Figure 2 showcases the relationship between various categorical features and car prices in the dataset. Cars with two and four doors exhibit similar median prices, though four-door cars display slightly higher variance, suggesting limited influence of the number of doors on pricing. Fuel type reveals a noticeable difference, with diesel cars having higher median prices than gasoline cars, reflecting their efficiency and performance. Turbocharged cars show higher median prices compared to those with standard aspiration, highlighting the added value of turbo engines. Among car body types, convertibles and hardtops command higher prices, while hatchbacks and wagons are positioned as more budget-friendly options. Engine type plays a significant role, with vehicles featuring ohcv and dohc engines priced higher due to their advanced technology, compared to the lower-priced ohc and l engines. Finally, the fuel system also impacts pricing, as idi and spdi systems are associated with higher prices, indicating their superior efficiency, while mpfi and 2bbl systems are linked to more affordable vehicles. The plot highlights the variations and outliers across these categorical features, providing insights into their influence on car pricing.

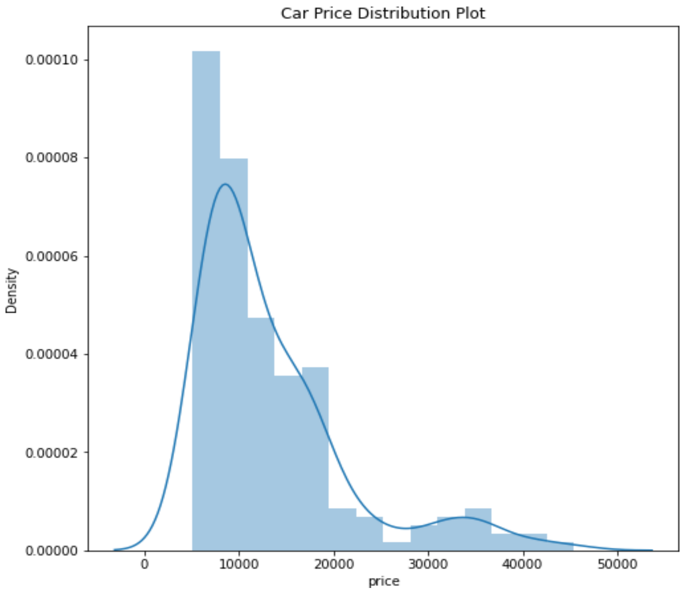

Figure 1: Distribution of Car Prices: Most vehicles are priced in the lower range, with a concentration around $10,000, while higher-priced cars are less frequent, showcasing a positively skewed distribution.

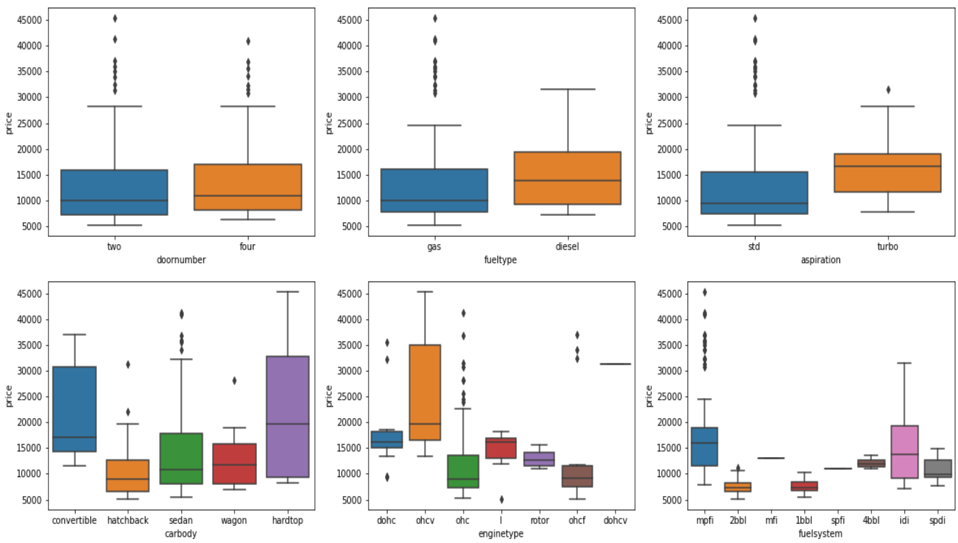

Figure 2: Impact of categorical features on car prices with variations and outliers via box plot

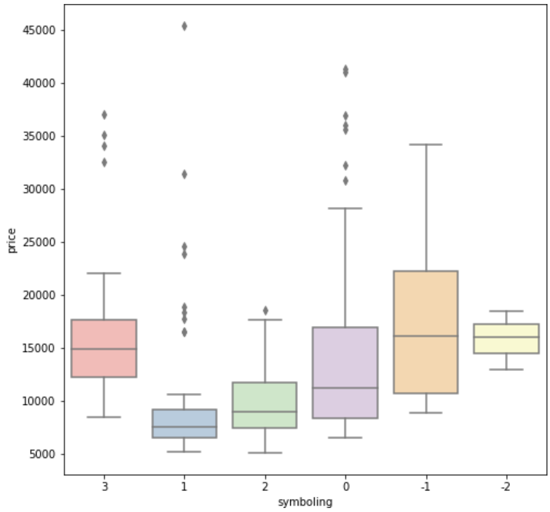

Figure 3: Impact of categorical features on car prices with variations and outliers via box plot.

The Figure 3 illustrates the relationship between car prices and their symboling values, which represent risk factors associated with the vehicles. Symboling is a process used by actuaries to assess and adjust the risk rating of a car based on various factors, with values ranging from +3 (high risk) to -3 (low risk). From the plot, cars with higher risk ratings (+3) generally have lower median prices, reflecting their riskier nature and possibly reduced desirability or features. Conversely, vehicles with lower risk ratings (-1 and -2) tend to have higher median prices, indicating that safer cars command a premium in the market. The variability in prices increases with lower risk ratings, suggesting that safety is a significant factor in pricing, particularly for higher-end models. Outliers are present across most categories, indicating exceptions where risky cars are priced higher or safe cars are priced lower, likely due to other influencing factors such as brand or specifications.

The heatmap (Figure 4) reveals significant correlations between various features in the dataset, providing insights into their relationships. Wheelbase shows a strong positive cor- relation with carlength, carwidth, and curbweight, indicating that larger vehicles tend to have greater dimensions and heavier weights. Similarly, carlength is highly correlated with curbweight, suggesting that longer cars are typically heavier, while also showing a negative correlation with highwaympg, implying reduced fuel efficiency for longer vehicles. Car- width is positively correlated with curbweight and enginesize, highlighting that wider cars tend to weigh more and have larger engines. Enginesize also has a strong positive corre- lation with horsepower, reflecting the direct relationship between engine size and power output.

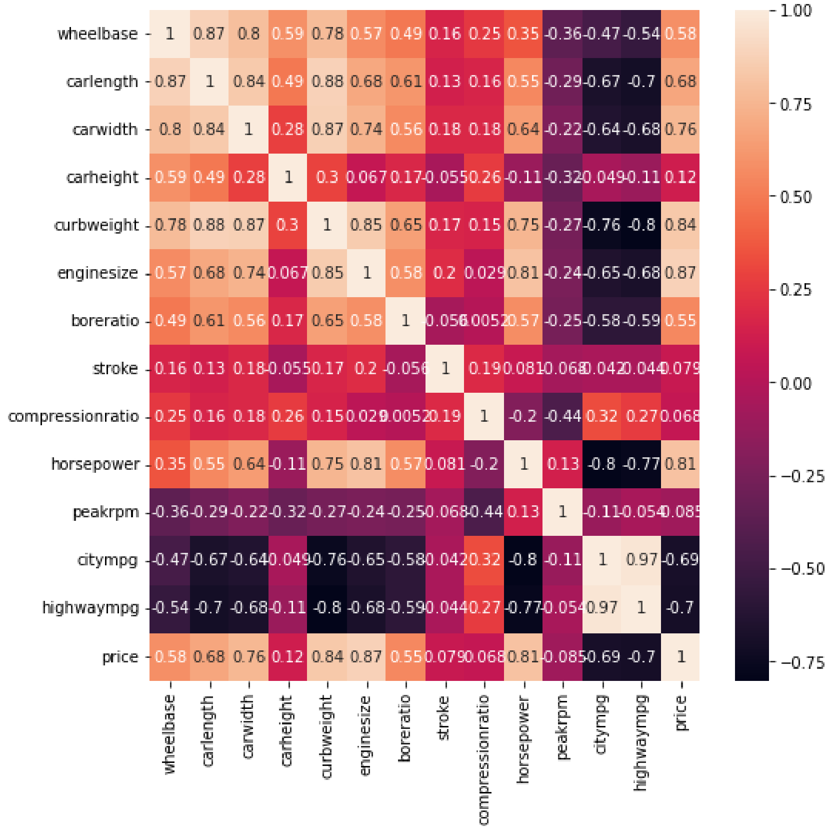

Figure 4: Impact of categorical features on car prices with variations and outliers via box plot.

Additionally, curbweight is highly correlated with enginesize and horsepower, further linking heavier cars with larger engines and higher performance, while showing a negative correlation with highwaympg, indicating reduced efficiency for heavier vehi- cles. Horsepower negatively correlates with both citympg and highwaympg, suggesting that more powerful engines often lead to lower fuel efficiency. Lastly, citympg and high- waympg are highly positively correlated, as expected, given that fuel efficiency tends to follow similar trends in both conditions. These correlations provide a deeper understand- ing of how vehicle dimensions, weight, engine characteristics, and fuel efficiency interact.

3.3 Methodology

PyCaret1, a low-code machine learning library, simplifies the process of performing regres- sion tasks by automating the workflow from data preprocessing to model evaluation. In this study, PyCaret was utilized to predict car prices, treating the target variable, price, as a re- gression problem. The original dataset consisted of 205 rows and 25 columns, which was transformed into a more concise dataset with 5 features after preprocessing. The data was split into training and test sets, with shapes of 143 rows and 62 rows, respectively. The pre- processing pipeline included imputation of missing values (mean for numeric and mode for categorical), normalization using the z-score method, and one-hot encoding for categorical features with a maximum limit of 25 encoded categories. Feature selection was enabled using the classic method with the LightGBM estimator, retaining the most significant fea- tures based on their importance. A 10-fold cross-validation approach was employed using KFold as the fold generator to ensure robust model evaluation. This comprehensive setup, combined with PyCaret's automated workflow, allowed for efficient exploration of multiple regression models to determine the factors influencing car prices.

The regression task using PyCaret follows a structured workflow to ensure a stream- lined and efficient process. The experiment begins with data preprocessing, where missing values are imputed (mean for numeric and mode for categorical variables), and features are normalized using the z-score method. Categorical features are encoded through one-hot encoding, with a limit of 25 encoded categories. Feature selection is then performed using the classic method with LightGBM as the estimator, retaining the most relevant features. Following preprocessing, the dataset is split into training (143 samples) and testing (62 samples) subsets. A 10-fold cross-validation approach, utilizing KFold as the fold genera- tor, is applied to ensure robust evaluation of model performance. PyCaret then trains and compares multiple regression models, automatically tuning hyperparameters and select- ing the best-performing model (Figure 5 and Figure 6). This systematic workflow ensures accurate predictions and insights into the factors influencing car prices.

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
et	Extra Trees Regressor	1935.4398	8477574.3131	2806.7468	0.8336	0.1694	0.1334	0.3520
catboost	CatBoost Regressor	2030.4826	8632348.8747	2854.3954	0.8479	0.1720	0.1379	0.5630
gbr	Gradient Boosting Regressor	2034.6540	9053669.1762	2879.5526	0.8156	0.1756	0.1427	0.2470
rf	Random Forest Regressor	2044.5456	9285507.5741	2961.5877	0.8087	0.1755	0.1417	0.3730
xgboost	Extreme Gradient Boosting	2212.0443	11206641.4500	3209.3432	0.7728	0.1904	0.1533	6.3760
dt	Decision Tree Regressor	2262.1776	11001294.6954	3185.1785	0.7855	0.2041	0.1617	0.2200
knn	K Neighbors Regressor	2385.3670	13607950.6000	3496.0129	0.7777	0.2099	0.1614	0.2220
ada	AdaBoost Regressor	2455.4455	10738975.6345	3197.1163	0.7777	0.2168	0.1954	0.3160
lightgbm	Light Gradient Boosting Machine	2523.0080	13142218.6876	3504.2675	0.7520	0.2164	0.1758	0.2090
par	Passive Aggressive Regressor	2941.7243	18233571.1244	4130.0865	0.7094	0.2737	0.2018	0.2020
huber	Huber Regressor	2989.2200	18604965.4558	4192.2492	0.6962	0.2853	0.2077	0.2080
br	Bayesian Ridge	3009.3737	17832740.5022	4134.9486	0.6821	0.3049	0.2203	0.1990
ridge	Ridge Regression	3014.9509	17906627.9581	4143.4318	0.6797	0.3032	0.2209	0.2110
llar	Lasso Least Angle Regression	3018.1568	17961944.6400	4150.2299	0.6778	0.3009	0.2212	0.2060
lasso	Lasso Regression	3019.3415	17984795.9766	4152.0721	0.6773	0.3013	0.2213	0.2600
lr	Linear Regression	3019.4599	17987278.3992	4152.2837	0.6773	0.3014	0.2214	0.7660
lar	Least Angle Regression	3019.4599	17987278.3992	4152.2837	0.6773	0.3014	0.2214	0.2010
en	Elastic Net	3024.3623	18711164.8428	4180.9756	0.7038	0.2622	0.2198	0.1980
omp	Orthogonal Matching Pursuit	3364.6704	21552909.8515	4579.8701	0.6071	0.3292	0.2504	0.1970
dummy	Dummy Regressor	6581.7186	75228056.2000	8429.7447	-0.1080	0.5461	0.5491	0.1990

Figure 5: Comparison of regression models, showcasing Extra Trees Regressor as the best-performing model with the lowest MAE, MSE, RMSE, and highest R2.

The input dataset was split into training, and testing sets following a 70:30 ratio to ensure balanced representation across the splits. This random division helped maintain a balance between training and testing data, avoiding overfitting and enhancing the model's generalization capability. Feature preprocessing, such as normalization and feature selec- tion, was performed prior to training to improve the model's learning performance. The use of k-fold cross-validation further ensured robust evaluation by testing the model's con- sistency across multiple data folds. Key features influencing the car price, such as curb weight, horsepower, and highway mileage, were retained during the feature selection pro- cess to maximize predictive accuracy.

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	2216.1018	9614373.3119	3100.7053	0.8515	0.1906	0.1796
1	1737.2880	4163203.7474	2040.3930	0.8226	0.1580	0.1396
2	3051.1404	27250427.7218	5220.1942	0.5974	0.2413	0.1923
3	2100.0627	10849187.1776	3293.8104	0.8827	0.1783	0.1610
4	1782.1144	5402687.0290	2324.3681	0.9377	0.1753	0.1263
5	2553.5769	10502429.3479	3240.7452	0.7999	0.2252	0.1976
6	2108.2723	6235732.2859	2497.1448	0.9218	0.1977	0.1710
7	2265.6807	9411476.3424	3067.8130	0.6140	0.1959	0.1819
8	1246.8782	3508934.5749	1873.2150	0.9497	0.1480	0.1137
9	2810.7648	12639967.2943	3555.2732	0.8889	0.1995	0.1600
Mean	2187.1880	9957841.8833	3021.3662	0.8266	0.1910	0.1623
Std	504.0829	6461459.9883	910.5976	0.1194	0.0268	0.0266

```
Fitting 10 folds for each of 10 candidates, totalling 100 fits
Original model was better than the tuned model, hence it will
be returned. NOTE: The display metrics are for the tuned model
(not the original one).
```

Figure 6: Summary of the evaluation metrics (MAE, MSE, RMSE, R2, RMSLE, MAPE) for 10-fold cross-validation, with Extra Trees Regressor achieving the best performance.

The model's performance was evaluated using the R2 metric, which measures the pro- portion of variance in the target variable explained by the features. The Extra Trees Regressor was found to be the most effective model, achieving a training R2 score of 0.999 and a testing R2 score of 0.857, indicating a strong fit. Additionally, the model had the lowest Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) compared to other models, demonstrating its robustness. The feature importance plot (Figure 7) revealed that ”curbweight,” ”horsepower,” and ”highwaympg” were the most significant contributors to price prediction. This analysis highlights the effectiveness of the Extra Trees Regressor in capturing the underlying data patterns and making accurate car price predictions while ensuring that only the most critical features were used (Figure 8).

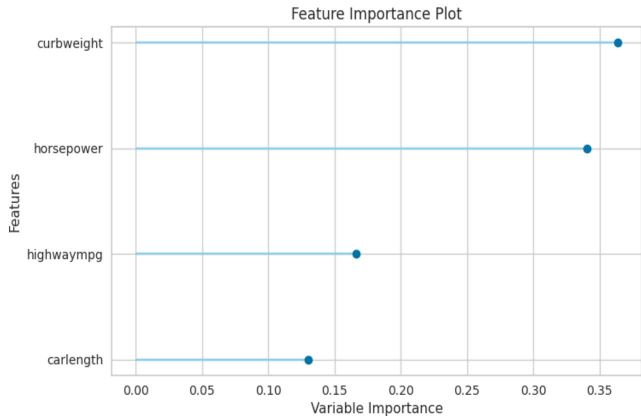

Figure 7: Feature importance plot highlighting the most influential features for predicting car prices, with 'curbweight,' 'horsepower,' and 'highwaympg' being the top contributors.

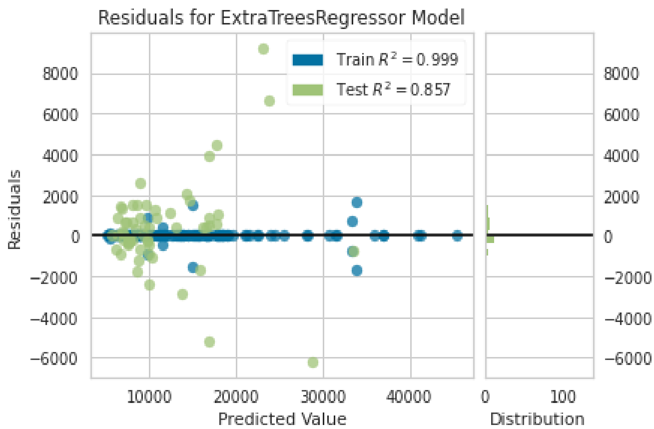

Figure 8: Residuals for the ExtraTreesRegressor model with Train R2 = 0.999 and Test R2= 0.857, illustrating model performance and prediction errors.

4. Conclusion

In this study, we employed PyCaret to explore and compare multiple machine learning models for predicting used car prices using a dataset sourced from the American automobile market. The comprehensive dataset included critical features such as vehicle specifications, engine details, and performance metrics. Through preprocessing—including imputation of missing values, normalization, encoding of categorical variables, and feature selection—we prepared the data for effective modeling.

The experiment followed a structured workflow, splitting the data into training (70%) and testing (30%) sets to ensure balanced representation and to prevent overfitting. Among the evaluated regression models, the Extra Trees Regressor emerged as the best performer, achieving a training R2 score of 99.9% and a testing R2 score of 85.7%. The model demonstrated robust predictive capabilities, effectively capturing the complex relationships within the data. Key features such as curbweight, horsepower, and highwaympg were identified as the most significant predictors influencing car prices.

The results underscore the effectiveness of using PyCaret for automating and stream- lining the machine learning pipeline in regression tasks. By leveraging automated feature selection and hyperparameter tuning, we were able to build a model that not only has high predictive accuracy but also provides insights into the factors that most significantly affect used car prices in the U.S. market.

5. Future Work

Building upon the promising results of the current model, future work can explore the integration of advanced techniques to further enhance prediction accuracy and model generalization. This could involve experimenting with ensemble methods beyond Extra Trees Regressor, such as gradient boosting machines or stacking multiple models to capture a wider range of patterns in the data. Additionally, incorporating fuzzy logic and genetic algorithms might offer improved handling of uncertainty and optimization in the prediction process.

Expanding the dataset to include more diverse and dynamic features—such as real-time market trends, consumer demand indicators, and economic factors—could provide a more holistic understanding of the variables influencing car prices. Furthermore, developing a fully automated, interactive system or application that utilizes the trained model would be beneficial for end-users. This system could serve as a recommendation engine, allowing users to input specific car attributes and receive accurate price estimates, thereby assisting buyers and sellers in making informed decisions. Future efforts will focus on enhancing this system's capabilities, user interface, and integration with a repository of used cars to provide personalized and up-to-date pricing information.

References

1. 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. Advances in Information and Communication Networks: Proceedings of the 2018 Future of Information and Communication Conference (FICC), Vol. 1, 413–422.
2. Zhang, X., Zhang, Z., & Qiu, C. (2017). Model of predicting the price range of used car. Awad, W. A., & ELseuofi, S. (2011). Machine learning methods for spam e-mail classification. International Journal of Computer Science & Information Technology (IJCSIT), 3(1), 173–184.
3. Durgesh, K. S., & Lekha, B. (2010). Data classification using support vector machine. Journal of theoretical and applied information technology, 12(1), 1–7.
4. Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. Int. J. Inf. Comput. Technol, 4(7), 753–764.
5. Noor, K., & Jan, S. (2017). Vehicle price prediction system using machine learning tech- niques. International Journal of Computer Applications, 167(9), 27–31.
6. Kuiper, S. (2008). Introduction to multiple regression: How much is your car worth? Jour- nal of Statistics Education, 16(3).