Predicting the price of used cars using data mining techniques

Seyyed amir mahdi Ghoreishi zadeh[1,*], Mostafa Setak[2]

[1]M.Sc. Student in Industrial Engineering Majoring In Macro Engineering systems, K. N. Toosi University

[2]Associate Professor, Department of Industrial Engineering, Economic and Social Systems, K. N. Toosi University

Abstract

In the automotive field, price analysis is very important for companies and individuals looking to gauge the market value of a vehicle before selling or buying. With the exponential increase in used car sales, dealers often take advantage of this trend by raising prices due to increased demand. The purpose of this research is to predict the price of used cars using supervised learning models of decision tree regression, random forest regression, AdaBoostRegressor to detect the effect of each feature on pricing and choose the best model, which finally among the tested models, random forest regression to The title of the top algorithm appears, which has the highest accuracy (94%) in predicting the price of used cars.

Keywords: Prediction, machine learning, used cars, regression algorithms, classification

1. Introduction

The used car market has continued to expand even as the new car market declined (Samruddhi et.al.,2020). However, predicting the price of used cars is an important and interesting problem. Accurate car price prediction requires specialized knowledge, because the price of products, especially products like cars, usually depends on many unique features and factors that require knowledge and expertise in this field.

Pricing is not only a science but an art that requires statistical and empirical formulas to create a profile for both the brand and the brand (Bukvić et.al., 2022). In recent years, due to global events and inflation, automobile-producing countries have consistently increased the prices of their products annually. As a result, many customers, instead of opting for bank loans or bearing the burden of hefty expenses to purchase a new car, choose to buy used vehicles. They consider functional cars at reasonable prices and make their purchases accordingly. According to a recent report on the Indian car market by Indian Blue Book, nearly 4 million used cars were bought and sold in 2018-19. The used car market has created this business for both buyers and sellers. Most people prefer to buy used cars because of the affordable price and they can resell them after a few years of use which may bring a profit (Samruddhi et.al.,2020). Also, according to the data obtained from the National Transportation Organization (Pudaruth, 2014), the number of registered cars between 2003 and 2013 has seen a significant increase of 234 percent. From 68 thousand 524 cars registered in 2012, this number has now reached 160 thousand 701 cars. With the tough economic conditions, there is a possibility of increasing the sales of imported second-hand (repaired) and used cars. Predicting the resale value of a car is not an easy task. It is common knowledge that the value of used cars depends on various factors.

The most important ones are usually the age of the car, its make (and model), the origin of the car (the original country of manufacture), the mileage (the number of kilometers it has traveled), and its horsepower. Due to the increase in fuel prices, fuel consumption is also very important. Unfortunately, in practice, most people do not know exactly how much fuel their car uses for each kilometer driven. Other factors such as the type of fuel used, interior style, braking system, acceleration, the volume of its cylinders (measured in cc), safety index, its size, number of doors, paint color, car weight, consumer reviews, prestigious awards obtained by the car manufacturer, its physical condition, whether it is a sports car, whether it has cruise control, automatic or manual transmission, whether it is owned by a person or a company, and other options such as air conditioning system, audio system, electric steering wheel, wheels Satellites, GPS navigation may all affect the price. Some of the special factors that buyers in Mauritius care about are the locality of the previous owners, whether the car has been involved in serious accidents, and whether it is a lady's car. As we can see, the price depends on many factors such as car type, fuel, color, model, mileage, transmission, engine, number of seats, etc. Unfortunately, information on all of these factors is not always available and the buyer must make a purchase decision based on only a few factors at a given price. In this work, we have considered only a small subset of the factors mentioned above. The article's structure is organized into distinct sections for clarity and coherence. It commences with a comprehensive literature review, delving into related works. The exploration begins with an overview of the comprehensive literature and then delves into related works. Following this, the third section introduces various learning methods and encompasses a preprocessing section. The fourth Section thoroughly examines different machine learning techniques and their applications, with a specific focus on their algorithms within the scope of this study. The next section, the fifth section, shows the obtained results and is dedicated to summarizing and concluding. Finally, the paper ends with a concluding section that summarizes the results and offers valuable suggestions for potential future research directions.

2. Literature Review

The realm of used car price prediction, although significant, is shrouded in a veil of limited research and scattered investigations, pointing to uncharted territories. This scarcity doesn't merely hint at gaps in knowledge but also underscores the untapped potential within this domain. The existing sparse research landscape, plagued by data constraints and fragmented studies, accentuates an overlooked yet promising area, inviting further exploration and in-depth investigation into the intricacies of forecasting used car values.

Bukvic et al Using data mining techniques conducted a study utilizing supervised machine learning to predict car prices based on the year of manufacture and mileage. Achieving an impressive 95% accuracy with linear regression, their findings highlighted that prices increase while mileage decreases, irrespective of the year of manufacture (Samruddhi et.al.,2020). Baharambe et al. concentrated on supervised machine learning techniques, exploring linear regression, lasso regression, and ridge regression. Employing Python libraries such as Numpy, Pandas, and Sklearn, they designed a GUI and evaluated machine learning accuracy. Among the algorithms tested, lasso regression exhibited the highest accuracy of 87.09%, outperforming linear and ridge regression methods (Bharambe et.al., 2022). Hankar et al. focused on supervised machine learning regression models, particularly multiple linear regression, to predict used car resale prices. Factors like mileage, fuel type, and year of manufacture were considered.

Their study identified gradient boosting regressor as a superior performer, displaying high R-squared scores and low root mean square error (Samruddhi et.al.,2020). Samruddhi et al. introduced a supervised machine learning model employing KNN regression to evaluate used car prices using the Kaggle dataset. Their experimentation with different trained test ratios yielded an accuracy of approximately 85%, establishing it as an optimization model for price prediction (Hankar et.al., 2022). Highlighted the need for more effective methods in used car price prediction, mentioning emerging fields like Artificial Neural Networks (ANN), Fuzzy Logic Systems (FLS), and Evolutionary Algorithms (EA) as potential solutions for computational intelligence challenges (Moayedi et.al., 2019 ; Nilashi et.al., 2018 ; Dreżewski et.al., 2018).

Enis Gejic et al. proposed an ensemble model utilizing various machine learning techniques (support vector machine, random forest, artificial neural network) to predict used car prices in Herzegovina and Bosnia. They attained an accuracy of 87% using data collected from the web portal www.autopijaca.ba (Gegic et.al., 2019). Pattabiraman Venkatasubbu and Mukkesh Ganesh constructed a statistical model to estimate used car prices based on previous customer data. They analyzed algorithms like lasso, multiple regression, and regression trees, aiming for improved prediction accuracy (Venkatasubbu& Ganesh, 2019). Yang et al. proposed a model for car price prediction based solely on product images, employing a custom convolutional neural network (CNN) architecture (Yang et.al., 2018). Nitis Monburinon et al. proposed regression models for predicting used car prices using data from a German e-commerce site. Their model with gradient-boosted regression displayed superior performance compared to linear regression and random forest models, showcasing a mean absolute error (MAE) value of 0.28 (Monburinon et.al., 2018). Nabarun Pal, Dhanasekar Sundararaman, Priya Arora, Puneet Kohli, and Sai Sumanth Palakurthy utilized random forest as a supervised learning method to predict used car costs. Their model, after detailed exploratory data analysis, achieved a training accuracy of 95.82% and a test accuracy of 83.63% (Pal et.al., 2019). The authors utilized a multiple linear regression model to predict both new and used car prices, utilizing tabular data sets for their analysis (Noor & Jan, 2017).

As a result, this research attempts to fill the gaps left by previous works by exploring different algorithms and expanding the scope of analysis with extensive datasets. The goal is to provide a stronger and more comprehensive understanding of predictive models for the field, emphasizing the need for broader methodological exploration to increase the accuracy and reliability of predictions in the field. In our analysis of used car prices, we use three distinct models: Decision Tree Regression, Random Forest, and AdaBoost Regressor.
Our goal is twofold: first, to evaluate the performance of newer algorithms, measuring their predictive power against traditional approaches. In the second step, we aim to compare the accuracy of these methods, to evaluate the models that have better results for predicting the price of used cars.

3. Applied Methods of Machine Learning

Machine learning, an integral subset of artificial intelligence (AI), revolutionizes how machines comprehend data, enhance performance through past experiences, and predict outcomes autonomously.
This paradigm encompasses an array of algorithms that grapple with vast datasets, refining models and executing specific tasks based on training data.
Categorically, machine learning branches into four principal types, delineated by learning methods and approaches:

- **Supervised machine learning**
- **Unsupervised machine learning**
- **Semi-supervised machine learning**
- **Reinforcement learning**

- **Supervised Machine Learning**: This method employs labeled datasets to train models, mapping input variables to corresponding outputs. It's instrumental when historical data with labeled outcomes are available, allowing the algorithm to learn patterns and predict outcomes accurately.

- **Unsupervised Machine Learning**: Operating on unlabeled data, this method identifies hidden patterns, similarities, and differences within datasets without guidance. It's advantageous when exploring data without predefined labels or when seeking to categorize information based on inherent structures.

- **Semi-Supervised Machine Learning**: Utilizing a blend of labeled and unlabeled data, this approach optimizes available information by initially clustering similar data unsupervised and subsequently labeling unlabeled data. It's beneficial when limited labeled data is available, offering a compromise between supervised and unsupervised learning.
- **Reinforcement Learning**: Driven by feedback mechanisms, this method involves agents navigating environments, learning from experiences, and optimizing actions based on received rewards or penalties. It's advantageous in scenarios where trial-and-error learning is crucial, such as game-playing or autonomous systems.

3.1. Dataset

We use a dataset from Kaggle related to the used car market, which is specifically designed to predict used car prices. This dataset contains 6,000 data records and a large number of features that are necessary to accurately predict and classify used car price ranges.

3.2. Preprocessing

Preprocessing is the preparatory phase in data analysis where raw data undergoes transformation and cleaning to enhance its quality and suitability for analysis. It involves tasks such as handling missing values, scaling features, encoding categorical variables, and removing outliers. Preprocessing ensures that the data is refined, standardized, and free from inconsistencies, laying a robust foundation for accurate and reliable analytical outcomes. This comprehensive dataset provides an overview of various aspects of the automotive industry. Including important details such as vehicle make, geographic location, year of manufacture, and technical specifications such as mileage, engine power, and seating capacity, it provides an overview of the automotive landscape, which can be seen in Table 1.
However, it appears that inconsistent naming conventions for vehicle registration and model were observed between the training and test datasets, potentially hindering the unified application of predictive models. This variation in naming conventions may reflect different data sources or cataloging practices, which is a unique obstacle in ensuring the generalizability of the model to unseen data. For this reason, we created a new attribute named "car" that contains only the name and model of the car, and We can see the new dataset after processing in Table 2.

Table 1. Sample data before data preprocessing

Unnamed	Name	Location	Year	Kilometers Driven	Fuel Type	Transmission	Owner Type	Mileage	Engine	Power	Seats	New Price	Price
1	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	First	21.56 kmpl	1462 CC	103.25 bhp	5.0	10.65 Lakh	9.95
2	Honda City 1.5 V AT Sunroof	Kolkata	2012	60000	Petrol	Automatic	First	16.8 kmpl	1497 CC	116.3 bhp	5.0	NaN	4.49
3	Maruti Swift VDI BSIV	Jaipur	2015	64424	Diesel	Manual	First	25.2 kmpl	1248 CC	74 bhp	5.0	NaN	5.60
4	Land Rover Range Rover 2.2L Pure	Delhi	2014	72000	Diesel	Automatic	First	12.7 kmpl	2179 CC	187.7 bhp	5.0	NaN	27.00
5	Land Rover Freelander 2 TD4 SE	Pune	2012	85000	Diesel	Automatic	Second	0.0 kmpl	2179 CC	115 bhp	5.0	NaN	17.50

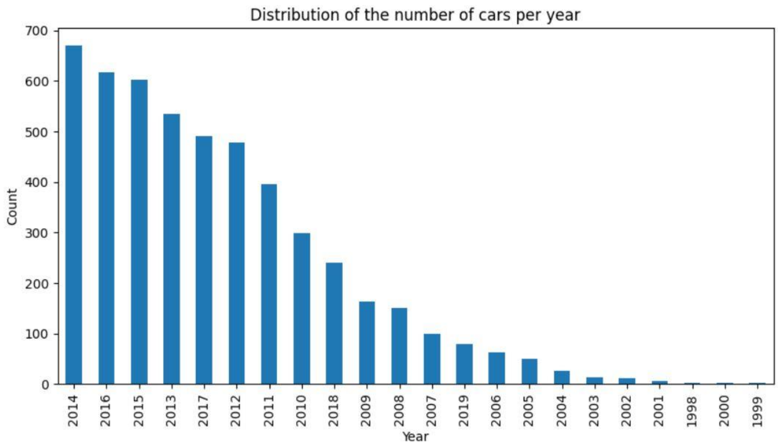
Table 2.Sample data after data preprocessing

Cars	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	Price
Maruti Ciaz	Kochi	2018	25692	Petrol	Manual	First	21.56	1462.0	103.25	5.0	9.95
Honda City	Kolkata	2012	60000	Petrol	Automatic	First	16.80	1497.0	116.30	5.0	4.49
Maruti Swift	Jaipur	2015	64424	Diesel	Manual	First	25.20	1248.0	74.00	5.0	5.60

Land Rover	Delhi	2014	72000	Diesel	Automatic	First	12.70	2179.0	187.70	5.0	27.00
Land Rover	Pune	2012	85000	Diesel	Automatic	Second	0.00	2179.0	115.00	5.0	17.50

Figure 1 shows the year of production of cars in the studied data set, in order of the largest number per year.

Figure1.Distribution of the year of production of cars of the second category of training data



3.2.1.Data Processing & Missing Identification

Upon closer examination of the dataset, it became evident that missing values, particularly in critical attributes like mileage, engine specifications, and power details, demand immediate attention. These attributes represent vital information pillars crucial for assessing vehicle performance and characteristics. The absence of these fundamental metrics not only jeopardizes accurate analysis but also compromises the integrity of machine learning models. A focused and systematic approach toward handling and resolving these missing components stands imperative, serving as the linchpin for data processing and preserving the dataset's fidelity, ultimately fortifying the predictive prowess of subsequent modeling endeavors.

3.2.2.Uniform characteristic unification & numeric conversion

To handle the different registration methods for car names and models in our dataset, it is necessary to create a "cars" attribute. The diversity of vehicle registrations and models in the dataset led to the integration of this new feature. Creating a "cars" feature harmonizes different car names, reduces common inconsistencies in datasets, and increases model generalizability. Detailed cleaning of "Milage", "Engine", "Power" and "Seats" attributes that convert textual data into numerical format ensures compatibility for robust analysis and modeling.
Presenting an overview of the cleaned datasets in Table 3 would be beneficial for a comprehensive understanding. This table can encapsulate essential statistics and characteristics, such as mean values, standard deviations, minimum and maximum values for numerical attributes, as well as count and unique values for categorical features. Additionally, including information on the data's size, including the number of rows and columns, and perhaps an indication of any remaining missing values or the successful removal of such instances, will offer a concise yet informative snapshot of the refined datasets. Table 3 can serve as a reference point for readers to grasp the fundamental attributes and statistical insights derived from the cleaned training and testing datasets.

Table 3. Summary statistics of cleaned training and testing datasets

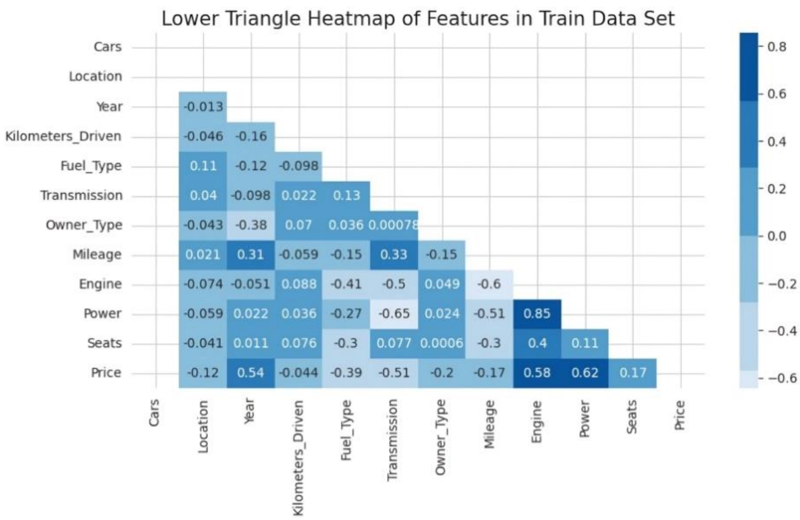
	Year	**Kilometers_Driven**	**Mileage**	**Engine**	**Power**	**Seats**	**Price**
count	4959.000000	4.959000e+03	4959.000000	4959.000000	4959.000000	4959.000000	4959.000000
mean	2013.373664	5.908272e+04	18.157413	1622.391813	113.386920	5.281508	9.521639

std	3.228135	9.934078e+04	4.482528	597.022078	52.908341	0.812495	11.154377
min	1998.000000	1.710000e+02	0.000000	624.000000	34.200000	0.000000	0.440000
25%	2011.000000	3.400000e+04	15.200000	1198.000000	78.000000	5.000000	3.500000
50%	2014.000000	5.302400e+04	18.160000	1493.000000	98.600000	5.000000	5.650000
75%	2016.000000	7.300000e+04	21.100000	1991.000000	138.100000	5.000000	10.000000
max	2019.000000	6.500000e+06	33.540000	5461.000000	550.000000	10.000000	160.000000

3.3.Exploratory data analysis (EDA)

Post-data preprocessing, a visual exploration ensues, unraveling nuanced facets of the used cars dataset. Employing an array of visualization techniques from bar charts and box plots to distribution graphs_we delve into the intricacies of each feature. This analytical journey aims not only to comprehend the data's inherent diversity but also to unravel the dynamics between different features, establishing insightful connections that extend to the target variable. Through this multifaceted approach, we gain a holistic understanding of the data, extracting valuable insights that pave the way for informed decision-making and strategic modeling. In a comprehensive analysis of the data set, we investigated the relationship between different characteristics and the price of cars. Correlation analysis showed that the three features with the highest correlation coefficients about car price were prominent: power, engine specification, and year of manufacture, which can be seen from Figure 2 of the heat graph.

Figure 2. Heatmap diagram of characteristics



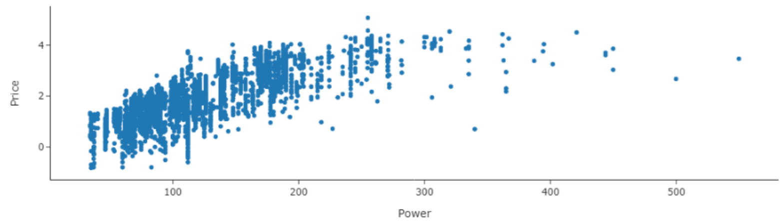
Power: Vehicle power, specifically classified by type, shows a significant relationship with price and can have a strong positive or negative correlation with their respective prices.

Engine Specifications: The engine specifications, encompassing factors such as horsepower, cylinder configuration, and other technical details, showed a substantial correlation with the price of the car. This suggests that the performance and technical capabilities of the engine play a crucial role in determining the overall value of the vehicle.

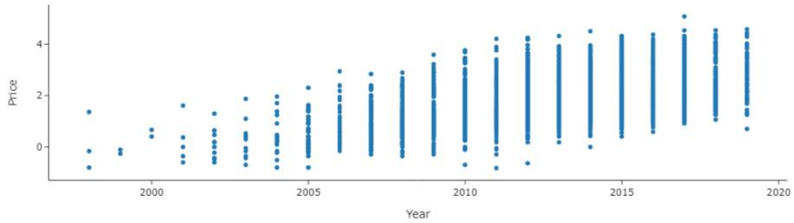
Year of Manufacture: The year in which the car was manufactured also exhibited a noteworthy correlation with its price. Generally, newer cars tended to command higher prices, reflecting the impact of depreciation, technological advancements, and overall market demand for more recent models.

In Figure 3 we use scatter plots to show the correlation between the target variable "price" of the car and the three key features that are most correlated with them. In these graphs, the vertical axis corresponds to the price and the horizontal axis corresponds to the power specifications, engine (automatic or manual), and the year of production of the car.

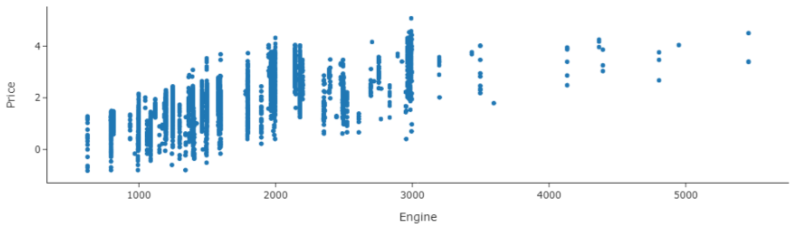
Figure 3.The relationship between price and the three main characteristics related to it



...



...



4. Machine learning algorithm used

4.1.decision tree

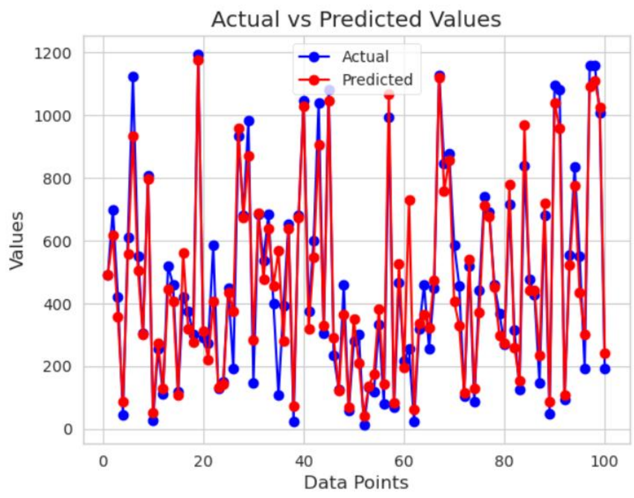
A decision tree is a predictive modeling tool used in machine learning and statistics for both classification and regression tasks. It works by recursively partitioning the dataset into subsets based on the most significant attributes, creating a tree-like structure of decisions.

At each node, the algorithm selects the feature that best separates the data, optimizing for criteria such as Gini impurity for classification or mean squared error for regression. This process continues until a specified stopping condition is met, resulting in a tree that can be used for making predictions on new, unseen data. The decision tree formula for regression can be shown in Equation 1 below:

$$\boldsymbol{y^{\wedge} = \sum_{m=1}^{M} C_m . I \quad (x \in R_m)} \qquad \textbf{(1)}$$

Here, $y^{\wedge}(x)$ is the predicted value for the input x, M is the number of terminal nodes (leaves) in the tree, C_m is the predicted value associated with the m leaf, and $I\ (x\ \in R_m)$ is indicator function that equals 1 if x falls into the m region R_m and 0 otherwise.
Decision trees are powerful tools for both understanding complex relationships in data and making predictions. They are particularly useful for tasks where feature interactions are important, and their interpretability makes them valuable for explaining model decisions to non-technical stakeholders. In Figure 4, the outcomes of car price prognostication utilizing the decision tree regression algorithm are depicted.

Figure 4. Prediction of decision tree regression model for 100 samples



4.2.Random forest

Random Forest is an ensemble learning technique widely used in predictive modeling for its robustness and accuracy. It's a collection of decision trees that operate together as a forest. Each tree within the forest works independently, generating predictions, and then combining these to yield an outcome. The algorithm follows a bagging method where multiple trees are constructed from subsets of the dataset, allowing each tree to learn from slightly different data.
This method's strength lies in its ability to reduce overfitting by combining various decision trees and averaging their predictions. Each tree in the forest gets trained on a random sample of data and random subsets of features, ensuring diverse and less correlated trees. The final prediction in a Random Forest is determined by aggregating the predictions of all trees, often taking the mean for regression or voting for classification. The algorithm's flexibility, efficiency in handling large datasets, and automatic feature selection through assessing variable importance make it popular for predicting car prices and other complex real-world problems.

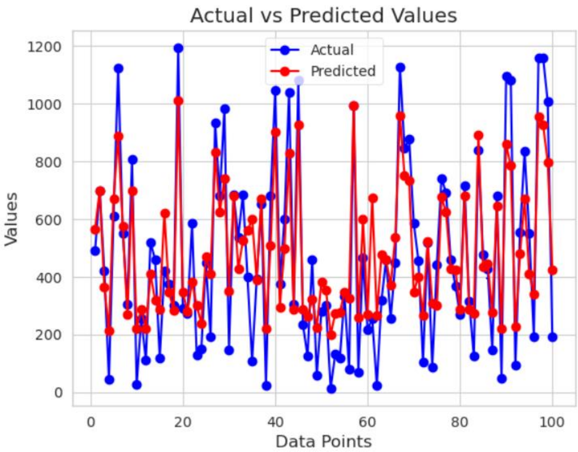
For a Random Forest, the prediction $y^{\wedge}$ for a given instance x is computed by taking the average (or mean) of predictions from individual trees as given in Equation 2 :

$$\boldsymbol{y^{\wedge} = \frac{1}{N}\sum_{i=1}^{N} y_i} \qquad (2)$$

Random Forest is an ideal choice for predicting second-hand car prices due to its ability to handle complex relationships among numerous car features. As second-hand car prices depend on various factors like mileage, age, make, and condition, Random Forest excels in handling these multifaceted interactions. The algorithm's strength lies in its capacity to manage high-dimensional data with numerous features, capturing nonlinear relationships and interactions between attributes effectively. With its ensemble of decision trees, each trained on different subsets of data, Random Forest can learn and adapt to diverse patterns within the dataset. Additionally, the method's robustness against overfitting ensures reliable predictions by aggregating multiple tree predictions. It can handle missing values, outliers, and noisy data gracefully, crucial in the diverse and dynamic landscape of second-hand car markets.

The Random Forest's automatic feature selection and ability to rank the importance of variables provide insights into which car attributes significantly influence pricing. Its versatility and efficiency make it a suitable model for accurately predicting second-hand car prices, considering the multitude of factors influencing their valuation. In Figure 5, the outcomes of car price prognostication utilizing the Random Forest regression algorithm are depicted.

Figure 5. Prediction Random forest regression model for 100 samples



4.3.AdaBoost Regression

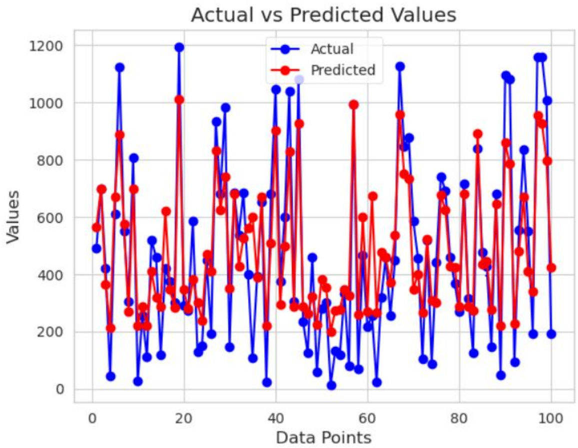
AdaBoost Regression is a cutting-edge algorithm used in price prediction, especially in the domain of second-hand cars. This method leverages a sequence of weak regression models, emphasizing the areas where the previous models fall short, ultimately creating a strong predictive model. The core idea behind AdaBoost lies in its iterative process. It trains a sequence of models sequentially, assigning more weight to the instances mispredicted by the preceding models. This way, each subsequent model focuses on the challenging instances, refining the overall prediction with each iteration. The final prediction is an amalgamation of the predictions made by each model, weighted by their performance. The model's ability to learn from its mistakes and constantly improve its predictions makes it highly effective in handling complex relationships within car features.
This formula includes assigning weight to individual models and collecting their predictions, which is mentioned in Equation 3 :

$$\mathbf{F}(\mathbf{x}) = \sum_{\mathbf{t}=\mathbf{1}}^{\mathbf{T}} \boldsymbol{\alpha}_{\mathbf{t}} \mathbf{f}_{\mathbf{t}}(\mathbf{x}) \qquad (\mathbf{3})$$

- **F(x)** is the final prediction.
- **t** represents each model in the sequence.
- $\boldsymbol{\alpha}_{\mathbf{t}}$ denotes the weight assigned to each model.
- $\mathbf{f}_{\mathbf{t}}(x)$ represents the prediction made by the respective model.

This adaptive boosting technique is particularly beneficial in capturing nuanced patterns and variations in second-hand car pricing, ensuring a more accurate prediction by iteratively learning from previous model mistakes. In Figure 6, the outcomes of car price prognostication utilizing the AdaBoost regression algorithm are depicted.

Figure 6. Prediction AdaBoost regression model for 100 samples



5. Results and discussion

In this study, we used three distinct algorithms for car price prediction: DecisionTreeRegressor, RandomForestRegressor, and AdaBoostRegressor. The first two, the decision tree and random forest algorithms, are known as standard and widely used approaches, while the AdaBoost algorithm was chosen for its innovative features. For evaluation, three key metrics were used: root mean square error (RMSE), accuracy on the training set, and accuracy on the test set. RMSE provides a measure of model prediction accuracy and quantifies the average deviation of predicted values from actual values. Accuracy on the training set assesses how well the model matches the training data, while accuracy on the test set measures its performance on previously unseen data and provides insight into generalization capabilities. The results of Table 4 show the performance measures for each model. AdaBoostRegressor showed an RMSE of 136.21, an accuracy of 83.28% on the training set, and 82.21% on the test set. DecisionTreeRegressor showed an RMSE of 112.60, achieving almost perfect accuracy on the training set (99.99%) and 87.84% accuracy on the test set. RandomForestRegressor performed better with an RMSE of 78.56 and an accuracy of 99.07% on the training set and 94.08% on the test set. As a result, the RandomForestRegressor model was identified as the superior prediction algorithm based on an extraordinary accuracy of 94%.References.

Table 4. Analysis of the final results of price forecasting

Rank	model	Root Mean Squared Error	Accuracy on Training set	Accuracy on Testing set
1	RandomForestRegressor	78.564588	0.990688	**0.940810**
2	DecisionTreeRegressor	112.599610	0.999993	0.878419
3	AdaBoostRegressor	136.206303	0.832809	0.822095

The results showed the distinctive features of each algorithm and highlighted the respective strengths and limitations. Without going into specific numerical details, it is clear that RandomForestRegressor emerged as the most promising model and exhibited superior accuracy in predicting used car prices.
In the ever-evolving landscape of big data and deep learning, future research could address the integration of broader datasets, potentially sourced from diverse sources. Investigating the use of advanced deep learning methods, such as neural networks, can further enhance the models' predictive capabilities. In addition, combining real-time market data and dynamically adapting models to changing market trends helps strengthen pricing forecasts in the automotive sector. Continuous refinement and exploration of advanced techniques remain at the forefront of accurate and adaptive car price forecasting.

Reference

[1] Samruddhi, K., & Kumar, R. A. (2020). Used car price prediction using k-nearest neighbor based model. Int. J. Innov. Res. Appl. Sci. Eng.(IJIRASE), 4, 629-632.

[2] Bukvić, L., Pašagić Škrinjar, J., Fratrović, T., & Abramović, B. (2022). Price Prediction and Classification of Used-Vehicles Using Supervised Machine Learning. Sustainability, 14(24), 17034.

[3] Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. Int. J. Inf.

[4] Bharambe, P. P., Bagul, B., Dandekar, S., & Ingle, P. (2022). Used Car Price Prediction using Different Machine Learning Algorithms. Int. J. Res. Appl. Sci. Eng. Technol, 10, 773-778.

[5] Hankar, M., Birjali, M., & Beni-Hssane, A. (2022, May). Used car price prediction using machine learning: A case study. In 2022 11th International Symposium on Signal, Image, Video and Communications (ISIVC) (pp. 1-4). IEEE.

[6] Moayedi, H., Mehrabi, M., Mosallanezhad, M., Rashid, A. S. A., & Pradhan, B. (2019). Modification of landslide susceptibility mapping using optimized PSO-ANN technique. Engineering with Computers, 35, 967-984.

[7] Nilashi, M., Cavallaro, F., Mardani, A., Zavadskas, E. K., Samad, S., & Ibrahim, O. (2018). Measuring country sustainability performance using ensembles of neuro-fuzzy technique. Sustainability, 10(8), 2707.

[8] Dreżewski, R., Dziuban, G., & Pająk, K. (2018). The bio-inspired optimization of trading strategies and its impact on the efficient market hypothesis and sustainable development strategies. Sustainability, 10(5), 1460.

[9] Gegic, E., Isakovic, B., Keco, D., Masetic, Z., & Kevric, J. (2019). Car price prediction using machine learning techniques. TEM Journal, 8(1), 113.

[10] Venkatasubbu, P., & Ganesh, M. (2019). Used cars price prediction using supervised learning techniques. Int. J. Eng. Adv. Technol.(IJEAT), 9(1S3).

[11] Yang, R. R., Chen, S., & Chou, E. (2018). AI blue book: vehicle price prediction using visual features. arXiv preprint arXiv:1803.11227.

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

[13] 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 Advances in Information and Communication Networks: Proceedings of the 2018 Future of Information and Communication Conference (FICC), Vol. 1 (pp. 413-422). Springer International Publishing.

[14] Noor, K., & Jan, S. (2017). Vehicle price prediction system using machine learning techniques. International Journal of Computer Applications, 167(9), 27-31.