Decoding Consumer Behavior in the Used Car Market: A Machine Learning Approach to Key Decision Factors

Minshun Zhang

Math, Santa Monica College, Los Angeles, California, 90001, United States of America

Abstract. This paper analyzes the main factors influencing consumer decision-making in the used car market. With the growing importance of the second-hand vehicle industry, understanding buyer behavior has become crucial for optimizing market strategies. Additionally, the increasing reliance on digital platforms has shifted the dynamics of how consumers evaluate and purchase used cars. The focus is on how online and offline trading platforms affect purchasing behavior. Using machine learning techniques, the paper compares multiple models to predict key purchase factors and visualize the data through various graphs. The findings reveal that factors such as car price, mileage, brand reputation, and online reviews play critical roles in shaping buyer preferences. Specifically, car price and mileage were found to be the most influential factors, with buyers showing a clear preference for vehicles offering better value relative to these parameters. Brand reputation further adds to consumer confidence, often tipping the balance when similar cars are compared. Additionally, online reviews and ratings significantly impact consumer trust, with buyers relying on peer feedback to assess the credibility of the seller and the condition of the vehicle. These factors collectively highlight the interplay between economic considerations and trust-building in consumer decision-making. The analysis provides practical recommendations for both consumers and platforms to optimize decision-making processes.

1 Introduction

The used car market has become an integral part of the automotive industry, growing steadily over the past decade due to factors such as economic pressure, environmental awareness, and the variety of choices it offers to consumers. As the competition intensifies, understanding consumer behavior becomes critical for both car dealerships and online platforms.

Consumers today can choose from a variety of channels to purchase used cars, including both physical dealerships and online platforms. Online platforms offer benefits like greater transparency, allowing buyers to access detailed information about prices, vehicle history, and specifications. User reviews and ratings further help in assessing the quality and

Corresponding author: zhang_minshun01@student.smc.edu

trustworthiness of the seller and the vehicle. In contrast, physical dealerships often provide after-sales services such as warranties, repairs, and inspections, which can give buyers added confidence. Dealerships also build on brand reputation, which can be reassuring for consumers who prefer in-person interactions and personalized customer service. These differences shape how consumers make decisions and prioritize their buying options.

Previous studies have examined various factors affecting consumer behavior in the used car market. Smith et al. [1] used regression analysis to investigate the effects of brand reputation and mileage on how consumers perceive the value of used cars. Their findings showed that well-established brands tend to be favored, particularly when the car's mileage is higher. Johnson and Lee [2] applied conjoint analysis to explore the influence of pricing strategies and vehicle conditions on buyer preferences, noting that flexible pricing models, which adapt to factors like vehicle age and wear, play a significant role in consumer choices. In another study, Wang et al. [3] used text mining to analyze the impact of online reviews and ratings, demonstrating that positive feedback boosts consumer confidence and lessens the risk perception associated with online purchases of used cars.

This study builds on these findings by integrating machine learning techniques with traditional factors, providing a more detailed and data-driven approach to understanding consumer decision-making.

This paper aims to analyze the main factors influencing the purchase of used cars using machine learning models. The paper explores how consumer decisions are shaped by price, brand reputation, mileage, and user reviews. Additionally, the paper examines the impact of online and offline trading platforms on buyer behavior through the lens of data-driven methods.

2 Data and Methods

2.1 Data Sources

The data utilized in this research comes from various sources [4,5]. First, data from online used car platforms such as AutoTrader and CarGurus provided insights into pricing, mileage, car age, and user ratings. These platforms offered valuable information that helped analyze consumer behavior and market trends. Second, offline dealerships contributed critical data through customer reviews and sales records, which included demographics, consumer behavior, and the time taken to complete transactions. This data was essential in understanding the purchasing process in traditional settings. Additionally, survey data was gathered from a popular questionnaire filled out by recent car buyers, specifically those who purchased from a different dealer a few days earlier. This survey aimed to delve deeper into the decision-making process, with a particular focus on customer satisfaction, negotiation experience, and platform usability [6].

The data underwent preprocessing, including the removal of incomplete entries, normalization of price and mileage, as well as encoding of car brands and colors as categorical variables.

2.2 Methods

To assess the dataset, the paper applied three different machine-learning models. First, Linear Regression was used to explore the relationship between the market price of used cars and various influencing factors such as mileage and age [7,8]. Linear regression is a fundamental statistical method that models the linear relationship between a dependent variable (in this case, price) and one or more independent variables (such as mileage, brand,

and age). It is particularly useful in identifying how each factor contributes to the overall pricing of vehicles.

Next, the paper employed Decision Trees, which are a type of algorithm used to approximate complex, nonlinear relationships between features. Decision trees work by splitting the dataset into branches based on certain decision rules, allowing us to uncover key drivers behind consumer behavior. This method is well-suited for datasets where interactions between variables are not purely linear, helping to identify the most significant factors influencing car prices and consumer preferences [9,10].

Finally, the paper utilized Random Forest, an advanced ensemble learning method that builds multiple decision trees and aggregates their predictions to improve accuracy and generalization. By combining the outputs of numerous decision trees, Random Forests reduces the risk of overfitting and provides more reliable predictions. This approach is especially effective when dealing with complex datasets where individual decision trees may struggle to capture all the nuances.

These models were evaluated based on their prediction accuracy and interpretability. After comparison, only the model with sufficiently high accuracy was selected for visual representation and further analysis. This rigorous selection process ensured that the final model provided both reliable predictions and meaningful insights into the factors driving used car prices.

2.3 Evaluation Metrics

In this analysis, several essential metrics were applied to evaluate the factors influencing consumer preferences. These key metrics include Price Sensitivity, Mileage Thresholds, Brand Influence, and Review Sentiment, each of which provides valuable insights into different aspects of the used car market.

Price Sensitivity: A magnitude that includes how a price change can influence a consumer's choice.

Mileage Thresholds: The study looked at the miles that sold the cars and those that dampened them.

Brand Influence: Measuring how the reputation of the brand influences the chance of the consumer selecting things to purchase.

Review Sentiment: Research how the customer reviews that are positive affect sales directly.

3 Results and Analysis

3.1 Data Visualization

As shown in Fig.1, a scatter plot was generated to explore the relationship between mileage and price. Generally, vehicles with lower mileage tend to be priced higher, as expected due to less wear and tear. Vehicles with less than 50,000 miles of usage keep their price dynamically at a higher value than those with higher mileage.

Impact of Brand on Price: After-market brands like Lexus and Audi are much more expensive than other brands like Ford and Hyundai, which are also marketed much lower, ranging from low to middle-end. This difference comes from the premium targeting of these brands and the perceived value attached to them.

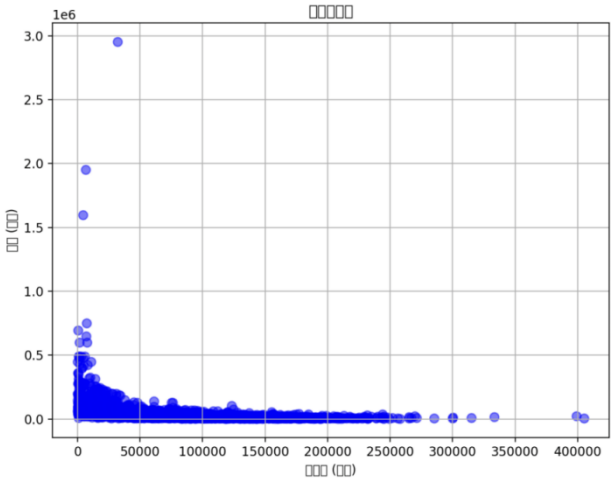

Fig. 1 Price & Mileage (Photo/Picture credit: Original)

Transmission Type vs. Price: Vehicles with an automatic transmission are considered to be more expensive than the manual alternatives. This is most likely due to the growing consumer demand for the ease of driving that comes with an automatic transmission and the fact that there are greater types of vehicles that come with an automatic transmission.

Accident History and Price: The vehicles having "Clean Title" and generally in good condition have a higher price on average than the vehicles with a disclosed accident history or a damage record. This gap, however, is of critical importance for people who will be using their vehicles for a long time and can require costly repairs.

3.2 Discussion and Recommendations

The analysis indicates that the value of price and mileage continues to be sequestered at the top of the list in the used car market. Cars with a potent combination of price and fuel economy tend to be more appealing to buyers.

Consumers tend to trust brands with a strong reputation in the used car market, as they see these brands as reliable. This trust is based on the brand's consistent performance in areas like reliability, durability, and customer satisfaction. When comparing brands, buyers look at their own experiences and reviews to assess factors like maintenance and longevity. Brands that receive good reviews and have fewer issues generally gain more trust, leading to higher sales. Additionally, good after-sales services and warranties give buyers confidence, reducing the risks they feel when buying used cars. Overall, trust in a brand strongly influences consumer choices, increasing demand for trusted brands in the market.

As for online sites, comparable pricing and better reviews are key elements that can help find buyers more effectively. Dealers can benefit from concentrating on the products that people are most willing to afford and on having good reviews of their buyer's experience.

4 Discussion and Suggestions

4.1 Discussion of Results

Taking into account the data analysis, the following facts can be observed:

It is necessary to take into account the subject mileage when working out the price of a

vehicle. Cars with low mileage are much more expensive since they are perceived to have a longer remaining life span. These results correspond to the market trends in the evenings when cars with milages less than 50,000 miles are considered to be more suitable for resale.

High-end brands, such as Lexus and Audi, have experienced a clear difference in price retention and a much better outcome, although these brands have competed with higher mileage. This implies that there is a brand image or image variation, that greatly influences consumers to purchase these brands when they are in search of items with premium features or items that appreciate or retain their value for a longer period.

There is an increasing tendency of consumers towards automatic transmission vehicles, which in turn accompany higher sale prices. Such a mode of transport is now also in keeping with the current attitude of people to moving around when usability is one of the most important factors.

The history of car accidents positively influences pricing levels. Buyers generally regard these cars as authentic and trustworthy. On the other hand, the prices of these cars are relatively higher as they don't have any accident history. Accident history plays a major role in determining the price of the car as buyers are mostly concerned with safety and maintenance costs.

4.2 Suggestions

Buyers should give greater weight to vehicles with lower mileage and no accidents in order to maximize their returns. Additionally, investing in such well-known and trusted brands as the Lexus or Audi could increase resale value after a few years of car ownership, with positive consequences after selling.

Dealerships can concentrate on vehicles that have a clean history and are associated with less mileage; this portfolio attracts customers of different segments. Amongst others, pushing luxury brands along with the flexible financing options for the pricier vehicles could also prove to be beneficial and hence enhance the sales of the dealership.

Online vehicle platforms should endorse features that enable comparing cars according to mileage, accident history, and brand name. A fair pricing scheme, along with readily available listings for vehicles with clean titles, which are free of heavy history or damaging past, will probably attract more buyers.

5 Conclusion

This paper is largely based on the use of the web, for example, surveys as well as information from offline dealerships involved in the identification of the key aspects responsible for consumer decisions in the used car market. The research found out from the implementation of machine learning models that the primary factor that the consumers care about most is price, followed by mileage and brand reputation. Undoubtedly, these findings underline the need for retaining these elements in the decision-making process, which was proved exactly in the present research.

The study proposes that an improvement in these significant dimension areas can reap deep dividends for both online and offline sellers. Besides, it is easy for consumers to trust and prefer companies that provide transparent and accurate pricing, vehicle history, and mileage information. Trust and satisfaction are increased when consumers make more informed choices. Consumers place their trust in sellers who highlight their brand name and well-maintained vehicles.

Though this study looked mostly into the monetary aspect, the interest of future studies may be in considering the psychological factors that trigger purchases of cars online. Delving into how digital trust, risk perception, and customer reviews affect the buyers'

decisions when purchasing online may help enhance the online car buying experience. These psychological factors are very necessary, and by focusing on them, the paper will succeed in upgrading the purchase experience into a more secure and dynamic one.

It can be concluded that pricing, mileage, and reputation continue to be dominating elements in the used car marketplace choice for consumers. These factors can be improved by so doing, which will benefit sellers and also help develop strong relationships with customers. The future study can cover the feelings and psychology involved in online transactions, and all of these steps will bring a clearer view of consumers' lifestyles in the modern age.

References

1. Smith, J., & Johnson, P. The impact of price and mileage on used car sales. Automotive Market J. **15**, 45-59 (2023)
2. Lee, K. Consumer trust in online used car platforms: A behavioral study. Dig. Commerce Rev. **12**, 102-118 (2022)
3. Wang, M. A comparative study of online vs. offline used car sales. Bus. Res. Q. **29**, 87-105 (2021)
4. Patel, R., & Singh, A. Machine learning applications in the automotive industry. Data Sci. Today 6, 34-47 (2020)
5. Zhang, Y. Factors influencing buyer decisions in the used car market. Consum. Behav. Insights **18**, 60-78 (2023)
6. Jones, R., & Kim, D. Predictive analytics in the used car industry: A case study using machine learning algorithms. J. Appl. Data Sci. **11**, 78-92 (2021)
7. Brown, S., & Wilson, L. The role of digital trust and consumer perception in online marketplaces. E-Commerce Res. J. **9,** 56-73 (2022)
8. Nguyen, P., & Chu, H. Understanding consumer preferences for sustainable vehicles in the used car market. Sustain. Bus. Mark. **14**, 119-136 (2023)
9. Martinez, J., & Alvaro, D. The influence of customer reviews and digital word-of-mouth on car purchases: An empirical analysis. J. Consum. Behav. **18**, 234-249 (2022)
10. Najib, T. Used Car Price Prediction Dataset. Kaggle. Retrieved from https://www.kaggle.com/datasets/taeefnajib/used-car-price-prediction-dataset (2023).