

Predictive Analysis on Used Cars

Background

Buying or selling a used car can be challenging as many factors contribute to deciding the correct resale prices. In recent years, the technology boom in the USA has disrupted the way people buy cars. New automobile companies are entering the market like Tesla. According to (Ifthikar & Vidanage, 2018), the demand for electric vehicles has forced manufacturers to switch to sustainable energy. This rush to become sustainable and be ahead in the market has led companies to produce new designs and features. It is indirectly affecting the value of used cars in the USA. Also, exceptional situations like the COVID-19 pandemic can affect the automobile industry, as described in the article by Zach Hope in detail (Hope, 2021). Even though the market is volatile due to indirect influences, used car prices still can be evaluated by some stable factors.

Relevant features for deciding used car value and how it influences the price

The factors which directly influence the prices are:

1. Aesthetics like colour and design- Paint jobs, stickers and modifications can negatively affect the value according to Cherise Threewitt in his article (Threewitt, 2020). Basic maintenance is enough to get a good price. Also, some car designs are very trendy and remain in fashion even after they are taken off the market.
2. Mechanical specifications- Number of engine cylinders, engine type (electric/gas/hybrid), engine displacement, odometer reading, type of transmission (manual/auto/semi-automatic), tank volume, horsepower, wheelbase, and alignment issues can make a big difference in the value of used cars.
3. Mileage- Cars used for driving long distances; mileage is a major factor that can save fuel. For instance, delivery drivers, taxi drivers consider the mileage as it is a huge expense for them.
4. Make- According to (Griffin, 2018), by the end of the first year of buying a car, the resale value of a car depreciates by 24 percent. The price keeps dropping every year and in three years, an average car can be almost half the price as compared to a new one.
5. Model/brand- Some car brands have an excellent resale value but some brands, in general, are not very good resellers. This is due to the low availability of parts or can be due to a faulty product by the company. However, some discontinued models become rare cars in demand like the Toyota CF Cruiser, as mentioned by Cherise Threewitt in his article (Threewitt, 2020).
6. Features- Extra keys, space, air conditioning, seating capacity are a few features that can make a difference in the resale price.
7. Condition of tyres- good quality tyres can cost a great deal. Worn out tyres will add the price of buying new tyres, as described in the article (A Look at Which Factors Can Reduce the Sale Value of Your Car| Buy My Car, n.d.)
8. Certified by the manufacturer- Certified cars come with company checks and are mostly in good condition. So, this raises the value of the car.

Some other aspects which can indirectly influence the prices of used cars are:

1. Hygiene levels inside the car.
2. Previous owner purpose of using car: commercial use.
3. Natural disasters: Pandemic, floods, hurricanes.
4. Location of trade- For instance, if the location of the car is in a scold area, it might not have issues with paint, or a convertible car might not sell in a cold area.
5. Accidental or not- accidental cars can have underlying issues.
6. Availability of maintenance records- Records can impact the value, as they can help resellers get a clear picture of the car maintenance history.

Methods used for price evaluation and assessment of used cars

Car evaluation is a subjective process that can be improved using an objective approach. There are two main ways to assess used car values:

1. Using an expert, unbiased opinion in deciding the best value of a used car. The problem with this approach is that even if it is from an experienced assessor, it can be wrong, as an expert might not consider all the parameters for an evaluation.
2. Another way, as suggested by (Ifthikar & Vidanage, 2018) is real-time web scrapping, where we collect data from various online resources and predict the prices using machine learning. This technique is more effective as this technique considers all parameters to design the model to evaluate the price. Adding all the parameters in an assessment impacts the value of a used car. Further use of machine learning techniques like SVM (Support Vector Machines), Decision trees, Linear regression can improve the accuracy of a model.

Identifying the data sources for value assessment of used cars

I found two structured data sources on the internet for the extraction of used cars data:

1. Data world: <https://data.world/data-society/used-cars-data>
2. Kaggle: <https://www.kaggle.com/austinreese/craigslist-carstrucks-data>

There are other data sources through companies like Edmunds.com, kbb.com, cars.com, carmax.com, from where we can collect data (web scraping) and get quoted prices for customers. Some companies charge a membership fee for their data access, while others are open source. Open source data is readily available, but it is not as good as the data on paid websites, as it offers correct prices. Most of the available information on data world and Kaggle is in tabular form and can be downloaded as a CSV file, while used car companies display data in raw text. Raw data can be a hassle to process and collect; sometimes data can be in a local language that needs to translate into English. However, most people speak English in the USA, and companies prefer the English language in data. So, this issue is not relevant in the USA. All data sources have many issues: missing values, skewness, redundant information, non-essential parameters, and incorrect labels. So, accessing the correct data can be a challenging task.

Managing data issues of used cars

Missing values can be replaced by mean, zero or other statistical measures like max, mode, and min. We can also remove the values if they are less than 5% of the total number of values in a parameter. Overall, it depends on the business case whether we choose to remove it or replace it with some other value. For skewness, we can use the Log transform function or other transformation functions. Redundant values can be eliminated easily in any tool like Excel, PowerBI, or Rstudio. Non-essential parameters like the owner's name and age can be removed as they do not contribute to our analysis. Incorrect labels can be renamed to make the values consistent, like Ford and ford can be combined into one value.

Choosing variables to build predictive models to assess used cars

The variables which have a huge impact on the value of the used car should be added to the predictive model. We can select these variables based on online research through used car websites and some basic knowledge in the automobile industry. We can think from a buyer's perspective of what matters the most when buying a used car. The variables useful to build a predictive model for deciding the price of used cars are:

- Brand- Hyundai, Toyota, Ford: more reliable/ Subaru, Volvo: less reliable and low availability of parts
- Colour- Black, Grey, White – better price/ Red, Orange, Green- low resale price.
- Number of engine cylinders- 2, 4, 6, 8, 12 cylinders (4, 6, 8 are best).
- Engine type - electric/gas/hybrid (hybrid is best).
- Engine displacement- large is good.
- Odometer reading- less is better.
- Type of transmission -manual/auto/semi-automatic (auto is best).
- Tank volume- large is good.
- Horsepower- large is good.
- Alignment issues- If yes, then a low value.
- Mileage- higher mileage, better price.
- Features: Extra keys, space, air conditioning, seating capacity: the more, the better.
- Condition of tyres- New, old, worn-out (worn-out tyres fetch low money).
- Certified by the manufacturer- If yes, then value increases.
- Accidental or not- If yes, then less value.
- Location of trade.
- Availability of maintenance records.

Conclusion

Overall, deciding used car prices in the USA can be a challenging task. It requires a great deal of research, and how parameters impact the price. Although some parameters are essential, due to the nature of the work and fluctuation in the market, it is hard to decide the factors contributing to the prices of the used cars. Predictive modelling using machine learning is the best technique to predict the prices of used cars in this business scenario.

References

1. Ifthikar, A., & Vidanage, K. (2018). Valuation of Used Vehicles: A Computational Intelligence Approach. 2018 8th International Conference on Intelligent Systems, Modelling and Simulation (ISMS). Published. <https://doi.org/10.1109/isms.2018.00011>
2. Hope, Z. (2021, February 13). 'Never seen anything like it': The used car market's wild coronavirus ride. The Age. <https://www.theage.com.au/national/victoria/never-seen-anything-like-it-the-used-car-market-s-wild-coronavirus-ride-20210212-p57213.html>
3. Threewitt, C. (2020, December 17). 10 Factors That Affect Your Car's Resale Value. HowStuffWorks. <https://auto.howstuffworks.com/buying-selling/car-resale-value.htm>
4. Griffin, K. (2018, September 20). Use This Formula to Predict a New Car's Value in a Few Years' Time. LiveAbout. <https://www.liveabout.com/predicting-what-a-used-car-will-be-worth-3308561>
5. A look at which factors can reduce the sale value of your car | Buy My Car. (n.d.). <https://www.buymycar.com/>. Retrieved August 5, 2021, from <https://www.buymycar.com/sell-my-car/factors-which-reduce-the-value-of-a-car>