CAR RESALE VALUE PREDICTION

TEAM ID:PNT2022TMID34722

TEAM LEADER:DINESH SINGH .M

TEAM MEMBER:ABDULLAH ARSHAK .N

TEAM MEMBER:AAKASH SARAVANAN K.V

TEAM MEMBER:AMAR SINGH.U

OVERVIEW

The Pre-owned cars or so-called used cars have capacious markets across the globe.
Before acquiring a used car, the buyer should be able to decide whether the price affixed for the car is genuine. Several facets including mileage, year, model, make, run and many more are needed to be considered before getting a hold of any pre-owned car. Both the seller and
The buyer should have a fair deal. This paper presents a system that has been implemented to predict a fair price for any pre-owned car. The system works well to anticipate the price of used cars for the Mumbai region. Ensemble techniques in machine learning namely Random Forest Algorithm, extreme Gradient Boost are deployed to develop models that can predict an appropriate price for the used cars. The techniques are compared so as to determine an optimal one. Both the methods provided comparable performance wherein extreme Boost outperformed the random forest algorithm. The researchers applied three machine learning technologies namely, Artificial Intelligence (AI), Support Vector Machine (SVM) and Random Forest (RF) separately. Authors collected data through different web portals.

A combination of such developments created a significant upsurge in the demand for used cars. In addition, factors such as affordability, the availability of used cars, the hike in the need for personal mobility, and the emergence of various online players to organize the market have resulted in the growth of the market. For instance, in 2019, Ebay Inc. launched a new eBay Motors application to enhance the used car sale and purchase process online. Until recently, automobile manufacturers and dealers have mainly focused on their new vehicle business with the exclusion of used cars, often viewed as a by-product. However, the competition in the market and the threat of new entries have created a great extent of an upsurge in the used car dealership. Moreover, the added quality and reliability of used cars changed the consumer attitude and increased the sales of the used passenger cars. Investing in used car management has become one of the market's requirements characterized by a slimming margin, relentless competition, and demanding consumers. Technological advancements such as the development of the internet and the introduction of hybrid and electric vehicles have changed the buyer position in the market. Moreover, consumers are now knowledgeable about the vehicle, their residual value, quality finance charges, availability, the price applied, and sometimes, the profit margin that the seller makes in a

closing deal. This knowledge has changed the dynamics and managed to turn customer intelligence to their advantage. Resultantly, consumers are more inclined toward buying used cars nowadays. Some of the key factors including transparency and symmetry of the information among the consumers and buyers, the online sales channel growth, certified used vehicle programs, and the strong position of franchise dealers play a vital role in driving the market for used cars. Various leading companies have set up online and offline stores worldwide to offer seamless used car buying experiences. For instance, in September 2020, AutoNation Inc. expanded its pre-owned vehicle store and opened two new stores in the USA Denver market.

The organization also announced its goal to open 130 AutoNation Inc. stores in operation across the USA by 2026. In both developed and developing countries, the used to new vehicle ratio has increased in the last few years, accounting for the reasons stated earlier. In addition, franchised dealers with support from OEM involvement in certification and marketing programs, online inventory pooling, and access to high-quality contracts are in a strong position to benefit from the growth in the market. The COVID-19 (Coronavirus Disease) pandemic has placed the automotive industry at great disruption. In the aftermath of the pandemic, the consumers are expected to prefer private conveyance. However, the financial disparities are expected to hamper the purchase of new vehicles and due to budget constraints, commuters are expected to opt for the used cars. Further, online, or digitally generated sales leads buy new vehicles in this pandemic period. Due to the pandemic, hybrids are expected to battle in the market for the next two or three years due to the current economic conditions.

Research on Second-hand Vehicle Evaluation System Based on Improved Replacement Cost Method

Adjustment of purchase price

In order to improve the authenticity of the second-car price assessment, it is necessary to readjust Advances in Engineering Research, volume 163 1291 the evaluation price of the second-car according to the vehicle's own market supply and demand, market competition, routine maintenance, and the technical condition of the vehicle itself. We use the analytic

hierarchy process to determine the weights of different factors, and adopt expert scoring to construct a judgment matrix.

Design of second-hand car valuation system based on improved replacement cost method

When designing the system, the second-car evaluation process should be analyzed first. And then the function module and database of the second-hand car evaluation system are designed, and finally the interface design of the system is completed.

Second-hand car identification assessment system hierarchy

In the logic design of the used car value evaluation system, the system's calculation work is dispersed in the application program. The designed system can complete the input of used car information, store the input in data form, and process the data according to the operation instructions of the improved replacement cost method. The system uses B/S structure. Presentation layer: It mainly implements the interaction between users and systems and consists of Web pages with different functions. The user presentation layer displays the system to different users. The user makes requests to the business logic layer through operations such as selection and input. The result of the process is returned to the presentation layer and displayed to the user. Business Logic Layer: This layer is the middle layer between the presentation layer and the data access layer and is the most critical part of the system. This section uses IIS on the Windows platform and the programming language uses ASP. This layer is composed of upload components and file management components. The upload component can upload the picture information of the used car; the file management component can add, delete, and perform other operations on the file. Data access layer: It mainly saves and reads data, and reads the information stored in the database to the business logic layer. Data reading is done by the ADO component, and the data processed by the business layer is saved in the database.

System Process Analysis

When evaluating a second-hand car, selecting the appropriate valuation model is necessary. According to the need to enter the evaluation parameters, different valuation models need to enter different parameters. According to the evaluation parameters input by the system, the

client verifies the validity of the entered form data through the data verification control. If the verification is valid, we should submit the entered form data to the server. The server generates an improved replacement cost method valuation formula to calculate vehicle prices based on formulas and parameters, and save the calculation result to the database, and return the calculation result to the client as well as.

Design of system function module

The second-hand car value assessment system is mainly used to evaluate the value of Advances in Engineering Research, volume 163 1292 second-hand cars and provide information on the valuation of used cars. The system mainly includes a system management module, a second hand car parameter management module, a second-hand car evaluation management module, and an evaluation information query module. System management module is mainly to achieve user management and system login. After the user is registered, the user name and password of the system are obtained. If the user name and password are correct, the user enters the system and checks related information. If you make a mistake, relevant tips should be made. The administrator of the system can manage the registered user information, including adding, deleting, etc., but cannot modify the related parameters of the system. The parameter management module is mainly used to input the relevant parameters of the second-hand car, calculate the new rate of the second-hand car according to the corresponding calculation formula, and display the related calculation result. The system administrator can add and retrieve values of the second-hand car's vehicle condition parameters and new rate options. The second-hand car evaluation management module is mainly to obtain the evaluation price of the second-hand car and to evaluate the sale price. Click Save to save the data information in the system database. The evaluation information query module stores vehicle information that has been evaluated. Through this module, the parameters of the second-hand car, the evaluation price, and the evaluation price of the second-hand car can be queried.

Vehicle Type Insights

The conventional vehicle segment accounted for a share of over 40.0%, in terms of shipment, in 2021. The electric vehicle segment is expected to register a significant CAGR over the forecast period, complemented by the hybrid vehicle. In the last few years, used electric

vehicle prices continue to remain viable for consumers, and this plays a significant driving factor for electric vehicle sales. According to the last few years' price analysis, used electric vehicles' prices have been lower than the used hybrid vehicles. Electric vehicle traits such as technology

driven performance, in the luxury vehicle segment, provides a status symbol and supports sustainability, thus creating a significant volume demand for used EVs. Conventional gasoline vehicles with large inventory offer multiple choices at an affordable price. This segment of vehicles accounted for the maximum share in all sizes, including compact cars, mid-size, and SUV cars. Further, growing concerns over climate change and increasing pollution have created a great demand for a substitute for conventional gasoline vehicles. Hence, there has been significant growth registered by the electric used cars in the market.

Fuel Type Insights

The petrol segment accounted for the largest volume share of over 40.0% share in 2021. This is attributed to the declining usage of diesel vehicles as the government discourages the purchase of used diesel vehicles. The other segment is expected to witness significant growth over the forecast period. In developing countries, CNG powered vehicles have also shown a sustainable upsurge in used vehicle volume sales. Emission standards for the positive ignition (gasoline, NG, LPG, ethanol) and compression ignition (diesel) vehicles have become one of the reasons for the slump in sales of diesel vehicles. Moreover, excessive emission of NOx by the diesel engine can be attributed to the decline in diesel engine vehicle sales and an increase in the substitute market. The petrol-fueled car emission standard is less stringent compared to diesel-fueled passenger cars. Furthermore, petrol cars with a refined engine, decent fuel efficiency, and strong top-end performance attracted a large consumer base in the last few years and are expected to continue with the same in the coming years. In addition, increasing inventory for petrol-based SUVs became one of the driving factors of the petrol segment.

Size Insights

The SUV size segment accounted for the largest volume share of over 35.0% in 2021. With the changing landscape in the automotive market, the SUVs segment has caused the downfall of other segments. Offering space and size while remaining compact compared to off-road vehicles, SUVs are considered ideal drives by buyers nowadays in various regions. With great demand and a wider supply network, residual value for SUVs is higher nowadays for the

market. The European region has witnessed significant demand traction for the used SUVs market. The compact size segment is expected to register a significant CAGR over the forecast period. This is attributed to people's preference for economical and compact size vehicles. Compact size vehicles with a high production rate and huge inventory have been preferred among the franchised owners. Easy availability with affordable prices fuelled the demand for the used compact vehicle in the last few years. However, with the changing consumer preferences and advancements in SUVs, the used SUVs have shown significant growth and this is expected to continue in the coming years.

Key Companies & Market Share Insights

The key players in the market are focusing on expanding the customer base to gain a competitive edge in the market. Thus, vendors are taking several strategic initiatives, such as collaborations, acquisitions & mergers, and partnerships. For instance, in 2020, Volkswagen announced a major investment in the market for used cars by a collaboration of its own used car chain, Das Welt Auto, with various used car platforms. Mainstream automakers have also been expanding their presence in this space with their pre-owned car sales networks like Maruti Suzuki's True Value, M&M Mahindra's First Choice Wheels, and Toyota's U Trust. Some prominent players in the global used car market include:

1. Alibaba.com
2. eBay Inc.
3. TrueCar, Inc.

METHODOLOGY:

There are four different methodologies used to predict the value of a resale car. The first method is the multiple linear regression analysis, the Pearson correlation coefficient (r) was computed between different pairs of features. The correlation between mileage and volume of cylinder, manufactured year and its mileage, mileage and price, cylinder volume and price are plotted in graph to obtain a linear graph. Also, the linear regression, logarithmic regression and regression coefficients are noted. Next K-Nearest Neighbours (kNN), is a machine learning

technique in which the new data is compared to all the existing records in order to locate the best match. Only three attributes were considered namely the make, year and cylinder volume. The data for year and cylinder had to be normalize to prevent large values from over-shadowing smaller values. Performance starts to degrade for higher values of k. The next method is Decision Trees, here the prices were grouped into six nominal categories as most of the popular decision tree algorithms cannot handle numeric outputs. There are many gaps in the ranges that have been defined because there were no cars within these ranges although it is certainly possible to get new data which fits within these zones. The Random Forest algorithm is very good at classifying the data based on the whole training set only however when the data is split between a training set and a testing set. The last method is the Naïve Bayes, it is one of the most useful machine learning technique. It is very easy to implement in software and the accuracy is usually as good as more complex algorithms.

Machine Learning Algorithms

The data, after being cleaned and normalized, is split into training and test data using a randomized 80-20 split. This is to ensure that the data used for testing does not contain any of the data used for training. Thus 20% of the data is reserved for testing purposes (see 4.4 Inference). The training dataset was used to train the four price prediction ML models chosen: Multiple Linear Regression, Lasso Regression, Ridge Regression, and Random Forest Regression. All machine learning algorithms used in this report were imported from the s k learn library. Some models were provided input parameters to implement. The motivations for the choice of input parameters are explained in this section for the models that require them. The Ridge Regression model was implemented with the argument alpha=0.01. An assortment of different alpha values were tried, and lower values performed slightly better. Values lower than 0.01 didn't noticeably improve model accuracy. Lasso Regression was similarly implemented with an alpha=0.01 value after testing. Lower values gave better prediction results. Random Forest was implemented with default parameters and random state=0. The random state is necessary because it is a stochastic process that takes a seed value to begin.

When testing different values, there was no noticeable performance increase. The random state throughout training was therefore set consistently to 0, minimizing stochastic behavior resulting from varying the random state.

Inferred Price Plots

For the first iteration of inferences (both training and testing), scatter plots were created for each algorithm where each data point is the actual price, plotted against the inferred price. A line of best was then calculated and drawn through these points as well as a line showing the actual values, y=x. If the algorithms do not demonstrate any systematic error, the line of best fit should match this line. See Appendix A for these scatter plots.

Measurement of Average Simulated Depreciation

The average depreciation was shown to be the same for Linear Regression, Ridge Regression, and Lasso Regression, independent of the age of the vehicle. This value was approximately 9.7%. The Random Forest Regression algorithm showed approximately 13.2% depreciation for cars that were 2-4 years old, and approximately 14.6% for cars that were 13-15 years old.

Project method discussion

As described in the project method description, this work utilizes quantitative analysis with measurements of the implemented machine learning algorithms to achieve the purpose of this work and answer the research questions. To be able to take the performance measurements required as part of this, the machine learning algorithms to compare had to be implemented correctly and using a suitable dataset. The requirements for this dataset were for it to be large, be representative of the used car market which a consumer valuation tool is likely to target, span several years in its sales, and include many relevant features.

The project milestones were chosen to accomplish this. The project milestones for this research were all met. As part of Milestone 1, four machine learning algorithms were identified as the best candidates for comparison, based on previous research in similar areas 35 and which are commonly utilized for regression analysis. Milestone 2 was to choose an appropriate dataset. The dataset chosen met the requirements, although it would potentially have increased the performance of the ML algorithms if it had exceeded the requirements to a

greater degree.

Because of missing information in some of the features, most of the cars in the dataset were dropped as part of the data cleaning process. Additionally, more features that could be retained after the data cleaning process would be beneficial. Particularly the “zip code” feature, which gives information on the region had too many discrete values to be made continuous. A dataset with more localized sales and therefore fewer discrete values would be more useful. It is possible that the dataset is not representative of the used car market, that is to say that cars with certain features are more likely to be included in this dataset and overrepresented. Milestone 3 was to make appropriate normalizations to the data. This was accomplished but at the cost of reducing the size of the dataset and number of features. Milestone 4 was to implement each of the models chosen as part of Milestone 1. The data cleaning and normalization process was refined based on the requirement for the implementations. The same data was used for all of the models, and the data cleaning and normalizations made should be made to suit all of the models, not to increase the performance of any one model. This was to ensure fair comparison in the performance measurements. Milestone 5 was to make the performance measurements. The metrics for this were chosen to best answer the research questions. To evaluate performance, two metrics commonly used in ML and related to the loss function that each model optimizes during training were used. Additionally, emphasis was put on using a third metric that is more useful for evaluating use for a consumer tool, for several price categories of cars. Milestone 6 was to compare the predicted depreciation for each model. This was to answer research question 2. There are several alternative approaches to doing this, but the one used in this research was to average the percentage decline in predicted price for all cars of a certain year in the testing dataset, and for different ages to see how well the models predict geometric depreciation. Since it is unlikely that the same car was sold multiple times in the dataset, the true depreciation could not be known to evaluate the performance. Instead, the measured average depreciation was compared to an average for all cars obtained from previous research in this area. For this comparison to be valid, the dataset needs to contain years where the depreciation followed this typical average depreciation, 36 and the cars included in the dataset need to be representative of the total market. If these assumptions are true, then any deviation from the average can only be explained by the inability of the models to detect depreciation.

Conclusion

This article has designed the used car value evaluation system, which is based on the improved replacement cost method. The results show that the system application effect is better. Due to the short system design time, the system's function is not perfect. In the next step, the actual application effect of the system will be tested and evaluated. It is expected that the system will eventually meet the automotive evaluation requirements. In this paper, we have trained our model with used cars data set to predict the price. Here we have used the K nearest Neighbour algorithm and we got accuracy 85% where the accuracy of linear regression is 71%. The proposed model is also validated with 5 and 10 folds by using K Fold Method. The experimental analysis shows that the proposed model is fitted as the optimized model. In our future work, we will apply advanced machine learning techniques and validate the model with different methods to enhance the optimization of the model with improved accuracy.

References

[1]. Liu Lijuan, Liang Tianran. Replacement Cost Method——Discussion on the Appraisal Method of Used Car Identification.

[2]Tu Weixing. Research on Used Car Evaluation System[J]. Nanjing: Nanjing Forestry University, 2008

[3]PREDICTING THE PRICE OF USED CARS USING MACHINE LEARNING-SAMEERCHAND PUDARUTH, Computer Science and Engineering Department, University of Mauritius, Reduit, Mauritius.

[4] Chetna Longania, Sai Prasad Potharaju, Sandhya Deore, Dept of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India.