Car Price Prediction using Machine Learning

1   2   3   4

Mrs. T Veda Reddy, Y. Praneeth, Y. Sai Kiran, G. Sai Pavan

1 Assistant professor Department of Computer Science and Engineering

Anurag University, Hyderabad, Telangana, India

2, 3,4

Department of Computer Science and Engineering

Anurag Group of Institutions, Hyderabad, Telangana, India

Abstract – The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Indian Government in the form of taxes. So, customers buying a brand-new vehicle may be confident of the money they make investments to be worth. But, due to the increased prices of new cars and the financial incapability of the customers to buy them, used Car sales are on a global increase. Therefore, to find the car price which would be best suited for the buyer in India, we are going to predict its cost with the help of Machine Learning algorithms [1] which are made available by the Python Environment such as the Gradient Boosting algorithm. Our dataset comprises data related to different car brands with a set of parameters (Name, Location, Year, Fuel Type, Transmission, Owner Type, Mileage, Engine, Power, Seats, Price). The primary purpose is to design a model for a given dataset and predict the car price with better accuracy.

Keywords – car price, machine learning, Gradient Boosting.

Introduction

In this fast world, you don't have your own personal mode of transportation sort of an automobile, life will become even additional agitated. The public choose to obtain their automobile as a result of its convenience to commute between places, permits movement with an outsized cluster of individuals with fuel potency, and safe mode of transport. The used automobile marketplace is witnessing a boom in India, with the decision for luxurious vehicles sometimes increasing. Till a couple of years, owning a luxury automobile won't be a dream for varied shoppers, as a result of money hurdles, however, this is often bit by bit dynamic as shoppers can simply obtain used luxury vehicles. Machine Learning provides numerous ways through that it's easier to predict the worth of an automobile, by the previous information that is obtainable. We've enforced the model exploitation supervised Learning techniques of Machine Learning, which is outlined by its use of labeled information sets to coach algorithms to classify data or predict outcomes accurately. As the input file is fed into the model, it adjusts its weights till the model

has been fitted fittingly, which happens as a part of the cross-validation method. If there is also further transparency within the marketplace and fewer intermediaries, the seller ought to get the next value for a vehicle and therefore the shopper ought to get one at a lower fee as margins get reduced on every facet.

Motivation- India's used-car market is booming as consumers have a large variety of choices, straightforward funding, convenient digital sales channels, and a growing preference for private quality within the COVID-19 era [2]. costs of latest cars had been inflated considerably in past few years. and so, consumers have understood that a second user automotive could be a higher worth for cash instead. The pandemic has solely pushed that understanding more. So, deciding whether or not a second user automotive is well worth the denote worth once you see listings online will be tough. many factors, as well as mileage, model, year, etc., will influence the particular price of an automotive. Moreover, from the attitude of a trafficker, it's additionally a perplexity to cost a second user automotive befittingly.

Problem Definition - It is easy for any company to price their new cars based on the manufacturing and marketing cost it involves. But when it comes to a used car it is quite difficult to define a price because it involves it is influenced by various parameters like car brand, manufactured year and etc. The goal of our project is to predict the best price for a pre-owned car in the Indian market based on the previous data related to sold cars using machine learning.

Objective Of the Project - The goal of this project is to create an efficient and effective model that will be able to predict the price of a used car by using the Gradient Boosting algorithm with better accuracy.

Related Work

Some people preferred a good corridor, some are high or low. price with all of their demanded features, some are only weak for. notorious brands of the auto only. To elect the perfect auto is still a delicate task though some parameters like color, comfort, seating capacity, etc. are known [3]. That's why we tried to compare some algorithms for prognosticating auto buying purposes that which one gives better delicacy.

Execution of the Naive Bayes Classification method is proposed by Fitrina et. al [4]. Credulous Bayes is known as a simple probabilistic classifier. They connected this strategy for predicting buy. They utilized a dataset on 20 car buying information and got 75% precision. Srivastava et. al [5] connected the powerful learning strategy, Bolster Vector Machines(SVM) to different sorts of information like Diabetes Information, Heart Information, Satellite Data, and Carry Information. Those datasets have multi classes. They have too demonstrated the examination of the comparative consequences of the utilization of jumper's bit capacities on their paper.

A comparative examination of machine learning calculation was proposed by Ragupathy et. al [6]. In their paper, they tried to recognize and classify estimation, passed on in the fundamental text. They have collected their information from social media like Twitter, comments, web journal posts, news, status upgrades, etc. They also applied Credulous Bayes, Choice Tree, K-Nearest Neighbor, and Support Vector Machine classifiers for their

comparison purpose. Their objective was to discover the foremost effective classification technique and SVM came out with 72.7% which was the best accuracy. Another expectation framework utilizing administered machine learning strategy was proposed by Noor et. al [7]. They used multiple straight relapse strategies and anticipated vehicle prices. They got 98% precision on their framework. Buddy et. al [8] proposed a technique for anticipating utilized cars costs. In their paper, they utilized an Arbitrary Woodland classifier to anticipate the costs of used cars. To prepare the information, they made an Irregular Timberland with 500 Decision trees. At long last, they got 95.82% as preparing accuracy and 83.63% as testing precision. Pudaruth et. al [9] proposed another strategy for foreseeing utilized cars costs. In that paper, he connected numerous straight relapse investigations, k-nearest neighbors, Credulous Bayes, and Choice Tree which were used to make the forecasts. Osisanwo F.Y. et. al [10] proposed a Supervised machine learning method. They compared seven different Administered learning calculations and portrayed those. They moreover found the foremost viable classification algorithm established on the dataset.

A distinctive work on car buying was proposed by R.Busse et. al [11]. In their paper, they prioritized the psychological effect of climate. They connected projection predisposition and striking nature as two major mental mechanisms.

An unused imperfection classification procedure was proposed by veni et. al [12] to anticipate the course name of the" severity" tuple. Those information tuples were depicted by different trait-like Phase quality, Imperfection, Stage Abandoned, Affect, and Weight. They connected the Credulous Bayes classifier for forecast purposes. Jayakameswaraiah et. al [2] created an information mining system to analyze cars. They proposed a TkNN clustering calculation to predict the correct car. They moreover appeared the comparison of KNN and their proposed novel TkNN clustering. Another car price prediction procedure was proposed by Gegic et. al [13] where they utilized three machine learning procedures. They got 92.38% accuracy on the combination of all ML methods.

Another therapeutic work was proposed by Jabbar et. al [14] to anticipate heart malady in the determination framework. They utilized the K-nearest neighbor (KNN) calculation to anticipate it. The algorithm performs massively with 100% exactness. Peerun et. al [15] presented a strategy to anticipate rice of second-hand cars. In their paper, they utilized Manufactured Neural Systems. They applied it to a dataset of 200 records cars and compared diverse kinds of machine learning calculations. Yuan et. al [16] advertised a ponder on the forecast. He attempted to foresee the car deals based on some web look records. In show disdain toward these well-known works, there too exist some more challenging works. As a result, we center on the comparison of four sorts of well-known machine learning algorithms and attempt to discover which calculation gives the finest accuracy for our dataset.

Methodology

Data is collected from Kaggle.com [17]. The following attributes were captured for each car: Name, Location, Year, Fuel Type, Transmission, Owner Type, Mileage, Engine, Power, Seats, and Price expressed in Indian rupees.

After the data was collected and stored, the data preprocessing step was applied. The attributes with unexpected values are processed accordingly i.e. In our case we have replaced them with the most repeated value of the attribute. The cars without a price are discarded in prior.

To avoid conflict in mileage among different cars, all the mileages of cars are been scaled to a kmpl because most of the records are in km. To convert categorical data values into numeric attributes like (Company, Name, Location, Fuel, Transmission, Owner) we have used a one-hot encoding [18] approach.

Implementation and Evaluation

Gradient Boosting

Gradient boosting is one of the most powerful techniques for building predictive models. Gradient Boosting Algorithm [19] is generally used when we want to decrease the Bias error. Gradient Boosting Algorithm can be used in regression as well as classification problems. Gradient boosting involves three elements: A loss function to be optimized, A weak learner to make predictions, an additive model to add weak learners to minimize the loss function.

Before applying this model, we have used 70% of the data for training the model and 30% to test the model and evaluate its accuracy. Initially, we got a 77% accuracy score by applying GradientBoostingRegressor [20] to the data. After optimizing the parameters of the model by the GrindSearchCV [21] method, the model has produced an accuracy score of 92%.

Future Enhancements

A car price prediction has been a high-interest research area, as it requires noticeable effort and knowledge of the field expert. A considerable number of distinct attributes are examined for reliable and accurate predictions. The major step in the prediction process is the collection and pre-processing of the data. In this project, data was normalized and cleaned to avoid unnecessary noise for machine learning algorithms. Applying a single machine algorithm to the data set accuracy was less than 70%. Therefore, the ensemble of multiple machine learning algorithms has been proposed and this combination of ML methods gains an accuracy of 93%. This is a significant improvement compared to the single machine learning method approach. However, the drawback of the proposed system is that it consumes much more computational resources than a single machine learning algorithm. Although this system has achieved astonishing performance in the car price prediction problem, it can also be implemented using an advanced machine learning model and with Deep learning techniques to improve its efficiency and accuracy. Moreover, as innovation has been increased in automobiles and we can observe Electric vehicles have gained public attention and are preferred by most than a normal car.

Conclusion

Since India's used-car market is booming as buyers have a wide range of options, easy financing, convenient digital sales channels, and a growing preference for personal mobility in the COVID-19 era, car prediction can be a challenging task due to the high number of attributes that should be considered for accurate prediction. The main weakness of Gradient boosting is that it sacrifices intelligibility and interpretability. The main limitation of this study is the low number of records that have been used. In future work, we intend to collect more data related to electric vehicles [22] and combustion vehicles and to use more advanced techniques.

References

[1] Information regarding machine learning techniques and algorithms-https://en.wikipedia.org/wiki/Machine_learning.

[2] Coronavirus impact on used car sales-https://economictimes.indiatimes.com/topic/coronavirus-impact-on-used-car-sales.

[3] M. Jayakameswaraiah and S. Ramakrishna, "Development of data mining system to analyze cars using tknn clustering algorithm," International Journal of Advanced Research in Computer Engineering Technology, vol. 3, no. 7, 2014.

[4] F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah, and C. B. Harahap, "Implementation of naive Bayes classification method for predicting purchase," in 2018 6th International Conference on Cyber and IT Service Management (CITSM). IEEE, 2018, pp. 1–5.

[5] K. S. Durgesh and B. Lekha, "Data classification using support vector machine," Journal of theoretical and applied information technology, vol. 12, no. 1, pp. 1–7, 2010.

[6] R. Ragupathy and L. Phaneendra Maguluri, "Comparative analysis of machine learning algorithms on social media test," International Journal of Engineering and Technology(UAE), vol. 7, pp. 284–290, 03 2018.

[7] K. Noor and S. Jan, "Vehicle price prediction system using machine learning techniques," International Journal of Computer Applications, vol. 167, no. 9, pp. 27–31, 2017.

[8] N. Pal, P. Arora, P. Kohli, D. Sundararaman, and S. S. Palakurthy, "How much is my car worth? a methodology for predicting used cars prices using random forest," in Future of Information and Communication Conference. Springer, 2018, pp. 413–422.

[9] S. Pudaruth, "Predicting the price of used cars using machine learning techniques," Int. J. Inf. Comput. Technol, vol. 4, no. 7, pp. 753–764, 2014. 183 Authorized licensed use limited to: Carleton University. Downloaded on May 29,2021 at 09:56:13 UTC from IEEE Xplore. Restrictions apply.

[10] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: classification and comparison," International Journal of Computer Trends and Technology (IJCTT), vol. 48, no. 3, pp. 128–138, 2017.

[11] M. R. Busse, D. G. Pope, J. C. Pope, and J. Silva-Risso, "The psychological effect of weather on car purchases," The Quarterly Journal of Economics, vol. 130, no. 1, pp. 371–414, 2015.

[12] S. Veni and A. Srinivasan, "Defect classification using naive Bayes classification," International Journal of Applied Engineering Research, vol.

[13] no. 22, pp. 12 693–12 700, 2017. [12] E. Gegic, B. Isakovic, D. Keco, Z. Masetic, and J. Kevric, "Car price prediction using machine learning techniques," 2019.

[14]] M. Jabbar, "Prediction of heart disease using k-nearest neighbor and particle swarm optimization," Biomed. Res, vol. 28, no. 9, pp. 4154– 4158, 2017.

[15] M. C. Sorkun, "Secondhand car price estimation using artificial neural network."

[16] Q. Yuan, Y. Liu, G. Peng, and B. Lv, "A prediction study on the car sales based on web search data," in The International Conference on E-Business and E-Government (Index by EI), 2011, p. 5.

[17] Cars price dataset- https://www.kaggle.com/datasets/avikasliwal/used-cars-price-prediction?select=train-data.csv

[18] One hot encoding – "https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html".

[19] Gradient boosting –https://en.wikipedia.org/wiki/Gradient_boosting.

[20] gradient boosting regressor – https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html.

[21] gridsearchcv–https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.

[22] Electric vehicles analysis- https://www.iea.org/reports/electric-vehicles.