Web Application for Pre-Owned Automobile Price Estimation Using Machine Learning

Gunalan K [1], Sanjai B[2], Sanjay S U[3], Nitish J[4]

[1,2,3,4] *Computer Technology, Bannari Amman Institute of Technology, Tamilnadu, India.*

Abstract: *The manufacturer sets the price of a new car on the market, along with certain extra expenses paid by the government in the form of taxes. Customers may be confident that their new car purchase will be worth their money. Yet, sales of used automobiles are rising globally as a result of new car price increases and the financial instability of the consumers who purchase them. As a result, a Used Vehicle Price Prediction system with a recommendation system is required. This system efficiently assesses the value of the used car utilizing a number of factors. The current system involves a procedure where a vendor chooses a price at random and the buyer is unaware of the car and its current market value. In actuality, neither the seller nor the price at which he should sell the automobile has any notion of its current value. Also,the seller's prediction of the car's cost may not be more accurate, which could result ina loss. I proposed a model that will be very accurate to solve this issue. Regression algorithms like linear, lasso, and random forest are employed since their output is a continuous value rather than a categorized one. As a result, it will be possible to estimate a car's exact price rather than just its price range. A user interface that acceptsinput from any user and displays the car pricing in accordance with user inputs hasbeen installed.*
Key Word: Linear Regression, Lasso Regression, Random Forest, Price Prediction.

I.INTRODUCTION

The global market for used vehicles has a lot of promise. The price of a new car is established by the manufacturer, with the government incurring some additional costs in the form of taxes. Yet, due to the rising price of new cars and consumers' inability to buy new cars because of a lack of money, used car sales are increasing globally. In order to help consumers make informed judgements, our project aims to develop machine learning models that can accurately predict a used car's pricing based on its features.

Machine Learning is a field of neither technology booming with enormous abilities and applications in automating jobs,where neither human intervention is required nor explicit programming. The importance of ML is so great that we can see its applications trending everywhere in our day-to-day lives. ML has given solutions for many problems that existedbefore and have made businesses in the world progress to a new extent. To develop an efficient, effective and accurate model which predicts the used car's price according to the user's inputs?

Car price forecast project is the ability to forecast the cost of a car used by given various attributes (data) of that car. There Is a saying that a car loses 10% of its value the moment the user drives it off a lot. Given that the user would expect that one of the main predictors is the amount of kms driven in the car, since more kms of driving wears down thecar. Additionally, the user would expect the brand of the car to also be a factor in the used car price, since some brands of cars cost more and may be better made. The user expects to encounter some issues with multicollinearity since some aspects of cars may be highly correlated. For example, SUV cars will probably have larger engines and more doors. Larger engines are connected with more cylinders.

Machine Learning Concept:

Machine Learning is a field of neither technology booming with enormous abilities and applications in automating jobs, where neither human intervention is required nor explicit programming. The importance of ML is so great that we can see its applications trending everywhere in our day-to-day lives. ML has given solution for many problems that existed before and have made businesses in the world progress to a new extent. To develop an efficient and effective and accurate model this predicts the pre-owned automobile's price according to the user's inputs.

II.LITERATURE REVIEW

Author: Kiran, Description: This paper deals with the expected estimate for resale cost of a car is most significant in the field of present research and technology. Most significant attributes are considered for predicting the resale car value. The significant relationships among various attributes are found by the correlations establishments. Inthis research the price of the car is regarded as a dependent variable for target prediction. The prediction's input data was taken from the web. The suitability of random forest regression algorithm is identified and implemented In this academic study for accurately predicting the resale value of the vehicle based on most significant attributes those that have been chosen based on highest correlation. The outcome research demonstrates that the model's built-in accuracy is higher to 90 percent and error obtained is 10 percent.

Multiple Regression: A survey

Richardson in his thesis Richardson (2009), explained the work done on durable vehicles production by car producers which do not depreciate and long lifetime. It used a multiple regression analysis to demonstrate that hybrid cars (engine modification) are more valuable price than other vehicles. the reason for using hybrid engines is due to environmental concerns like global warming, climate change because of its higher fuel efficiency. They used factors like age of car, maker or brand of car, MPG (miles per gallon) and mileage for the thesis report study (Lodarosi, 2020; Nyandra et al., 2018).

Use of Supervised ML: A survey

Author: Prashant Gajera, Akshay Gondaliya, Jenish Kavathiya, Description: This paper deals with car price prediction with a platform that helps the people. The upcoming data with that platform which is made using machinelearning technology. Making use of supervised machine learning algorithms such 5 as linear regression, random forest regression. It helps build a statistical model which will be able to forecast the used car's price. For that, previous consumer data and a given set of features will help us. It will also compare the prediction accuracy of thesemodels to determine the unique one.

Author: Pattabiraman Venkatasubbu, Mukkesh Ganesh, And Description: This paper deals with the fact that the production of cars has been increasing in the past decade, with over 70 million passenger cars being produced in the year 2016. This has spawned the market for second-hand car prices. Which on its own has become a booming industry? The recent advent of online portals has facilitated the need for both the buyer and the seller to be better informed about the trends and patterns that determine the price of a used car in the market. With Machine Learning Algorithms such as Lasso Regression, Multiple Regression and Regression trees, we will try to develop a statistical model which will be able to predict used car's price, based on previous consumer data and a given set of features. We will also be comparing the prediction accuracy of these models to find out the unique one.

III .SYSTEM ANALYSIS

Demerits of the Existing System:

The information required at the cost assessment of the pre-owned vehicles is less in the current framework. Just variation, model, brand and the model year of the car and the assessed cost of the locomotive quoted by the vendor arethe main data infused in the majority of the current frameworks. In the current frameworks, the owners of the vehicles code the selling cost of the vehicles which is great for the proprietor's hand. The fulfilment of the purchaser and the dealer is less and the assessed cost isn't sensible. Also, the financier and representatives are associated with the current frameworks. In the current framework, mechanics are believed to give the last set of the vehicles after the assessment. Mileage and pull are ignored in the cost assessment information.

Merits of the Newly Proposed System:

The information required at the cost assessment of the pre-owned vehicles is more contrasted with the current framework. Other than the variation, model and brand of the locomotive, the mileage and Administration records of the automobile is likewise required for assessing cost for more prominent exactness. In the more seasoned or existing frameworks, the proprietors of the vehicles code the selling cost of the vehicles which is ideal for the proprietor's hand. The satisfaction of the both vendor and purchaser are more and no existence of financier and dealers required as an outside arbiter. Easy to use, more dependable, effectively reasonable, high precision in assessment of the cost price of trade-in vehicles. There is fewer requirements for a specialist to give a value set of the vehicle after examination. The more sensible and best resale cost is assessed through the recently proposed framework which satisfies both the vendors and purchaser's fulfilment.

IV.MODULE DESCRIPTION

Data Pre-Processing:

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning ideal. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not usually a case that we come across clean and formatted data. And while doing any operation with data, it is mandatoryto clean it and put it in a formatted way. So for this, the user can use data pre-processing task

Training:

A training model is a dataset that is utilized to train an ML algorithm. It consists of the sample output data and the corresponding sets of input data that rely on the output. The training model is utilized to run the input data through the algorithm to correlate the processed output against the sample output. The result from this correlation is utilized to modify the model.

Testing:

In machine learning, model testing is known as the process where the work of a fully trained model is evaluated on a testing set. This kind of ML testing is more similar to traditional testing. Users can write and run tests checking the work of the program. Applying the tests, users catch bugs in different items of the ML program. For the sake, users cantest that the hidden layers in a neural network are configured correctly.

Linear Regression:

       A machine learning algorithm called linear regression uses supervised learning as its foundation. The goal of linear regression is to predict the value of a dependent variable (y) based on an identified independent variable (x). So, x (the input) and y (the output) are found to be linearly related by this regression technique (output). It is frequently used to highlight the relationship between factors and forecasts.

Lasso Regression:

       A regularization method is lasso regression. For a more accurate forecast, it is preferred over regression techniques. Shrinkage is used in this model. Values in data that is shrunk tend to converge towards the mean, which is the centre. Simple, sparse models are motivated by the lasso approach.

Random Forest Regression:

       A supervised learning algorithm called Random Forest Regression employs ensemble learning techniques for regression. The ensemble learning method combines forecasts from different machine learning algorithms to produce predictions that are clearer than those from a single model.

Prediction:

       "Prediction" refers to the output of an algorithm .It has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome. Just like a hypothesis, a prediction is a type of guess. However, a prediction is an estimation made from observations.

V.PROJECT METHODOLOGY

There exist two initial phases in the system:

1. Training Phase
2. Testing Phase

Training Phase:

       The system is trained by using the data in the data set and fits a model in accordance with the algorithm choseneffectively.

Testing Phase:

       The framework is furnished with the sources of info and is tried for its operation. The exactness is validated. What's more, subsequently, the information that is utilized to prepare the model or test it, must be fitting. The framework is intended to identify and anticipate the cost of pre-owned vehicles and thus suitable calculations should be utilized to perform the two distinct errands. Earlier the calculations were chosen for additional utilization, various calculations were looked at for its exactness. The appropriate one for the undertaking was picked.

VI.RESULT AND DISCUSSION

       Metrics used to evaluate the algorithms in this paper are confusion matrix, classification report and accuracy score.

A confusion matrix has the total count of the accurately grouped occurrences along its cross and the count of the incorrectly classified instances in the rest of the matrix. We have used 4 class values; so, the matrix generated is a 4*4 matrix. A classification report gives the full report of the classification with parameters like recall, precision, f1-score, etc. Accuracy score gives the accuracy of the trained model after evaluating it using test data, for which we have sample 20% of the dataset The accuracy of the models can be improved by doing some data preprocessing steps likenormalization and standardization. Feature selection and extraction algorithms can be used to remove unsuitableand duplicative features to get better results. The same procedure used in this paper can be applied to predict the prices ofother products like cars, bikes, houses, etc. using the archival data containing features like cost, specifications, etc. Thiswould help organizations and consumers alike to make more educated decisions when it comes to price.

VII.CONCLUSION

       The expanded costs of new vehicles and the monetary inability of the clients to get them, Utilized Vehicle deals are on a worldwide increment. Subsequently, there is a pressing requirement for a Pre-owned vehicle Value Expectation framework which successfully decides the value of the vehicle utilizing different elements. The proposed framework will assist with deciding the precise cost of trade-in vehicle cost expectation. Vehicle Value Forecast was intended to get alternate pointsof view and ultimately contrasted their presentation and various models. Vehicle value expectation can be a moving undertaking because of the great number of traits that ought to be considered for the precise forecast. The significant stage inthe expectation cycle is assortment and pre-processing of the information. In this exploration, direct relapse and rope relapse, arbitrary woodland relapse were worked to standardize, normalize and clean information to keep away from pointless clamor for AI calculations. Information cleaning is one of the cycles that increments forecast execution.

References

[1] Agencija za statistiku BiH. (n.d.), retrieved from http://www.bhas.ba.[accessed July 18, 2018.]

[2] Listiani, M. (2009). Support vector regression analysis for price prediction in a car leasing application (Doctoral dissertation, Master

thesis, TU Hamburg Harburg).

[3] *Wu, J. D., Hsu, C. C., & Chen, H. C. (2009). An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(4), 7809-7817.*

[4] *Du, J., Xie, L., & Schroeder, S. (2009). Practice Prize Paper—PIN Optimal Distribution of Auction Vehicles System: Applying Price Forecasting,Elasticity Estimation, and Genetic Algorithms to Used-Vehicle Distribution. Marketing Science, 28(4), 637-644.*

[5] *Doan Van Thai, Luong Ngoc Son, Pham Vu Tien, Nguyen Nhat Anh, Nguyen Thi Ngoc Anh, "Prediction car prices using qualify qualitative data and knowledge-based system" (Hanoi National University)*