

USED CAR PRICE PREDICTION

V.Sravan kiran, assistant professor,M.tech, department of IT,sravan kiran_v@vnrvjiet.in

Rajath kala, B. Tech, Department of IT, rajath_k18@vnrvjiet.in

V. Nagesh, Btech, Department of IT, nageshv517@gmail.com

S. Navdeep, B. Tech, department of IT, navadeep1012@gmail.com

ABSTRACT: The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. 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, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and it's value in the present day scenario. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at. To overcome this problem we have developed a model which will be highly effective. Machine learning Algorithms are used because they provide us with continuous value as an output and not a categorized value. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.

Keywords- *Machine learning algorithms.*

1. INTRODUCTION

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models . We will compare the performance of various machine learning algorithms like Linear Regression, Ridge Regression, Lasso Regression, Elastic Net, Decision Tree Regressor and choose the best out of it. Depending on various parameters we will determine the price of the car. Regression Algorithms are used because they provide us with continuous value as an output and not a categorized value because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.

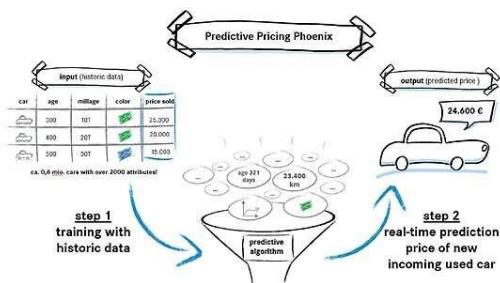


Fig.1: Example figure

The used car market is an ever-rising industry, which has almost doubled its market value in the last few years. The emergence of online portals such as CarDheko, Quikr, Carwale, Cars24, and many others has facilitated the need for both the customer and the seller to be better informed about the trends and patterns that determine the value of the used car in the market. Machine Learning algorithms can be used to predict the retail value of a car, based on a certain set of features. Different websites have different algorithms to generate the retail price of the used cars, and hence there isn't a unified algorithm for determining the price. By training statistical models for predicting the prices, one can easily get a rough estimate of the price without actually entering the details into the desired website. The main objective of this paper is to use three different prediction models to predict the retail price of a used car and compare their levels of accuracy. The data set used for the prediction models was created by Shonda Kuiper[1]. The data was collected from the 2005 Central Edition of the Kelly Blue Book and has 804 records of 2005 GM cars, whose retail prices have been calculated. The data set primarily comprises of categorical attributes along with two quantitative attributes.

Predicting the Price of Used Cars using Machine Learning Techniques

Abstract: In this paper, we investigate the application of supervised machine learning techniques to predict the price of used cars in Mauritius. The predictions are based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. The predictions are then evaluated and compared in order to find those which provide the best performances. A seemingly easy problem turned out to be indeed very difficult to resolve with high accuracy. All the four methods provided comparable performance. In the future, we intend to use more sophisticated algorithms to make the predictions.

car Price Prediction Using Machine Learning

Abstract: A car price prediction has been a highinterest research area, as it requires noticeable effort and knowledge of the field expert. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, we applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal autopijaca.ba using web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction

model was integrated into Java application. Furthermore, the model was evaluated using test data and the accuracy of 87.38% was obtained.

Price Evaluation Model In Second Hand Car System Based On BP Neural Network Theory

Abstract: With the rapid growth of the number of private cars and the development of the second-hand car market, second-hand cars have become the main choice when people buy cars. The online second-hand car platform provides both buyers and sellers the chance of online P2P trade. In such systems, the accuracy of second-hand car price evaluation largely determines whether the seller and the buyer can get more efficient trading experience. In this paper, the price evaluation model based on big data analysis is proposed, which takes advantage of widely circulated vehicle data and a large number of vehicle transaction data to analyze the price data for each type of vehicles by using the optimized BP neural network algorithm. It aims to establish a second-hand car price evaluation model to get the price that best matches the car. In this paper, the optimized BP neural network algorithm is used to select the optimal number of hidden neurons in BP neural network, which improves the convergence speed of the network topology and the accuracy of the prediction model. Through the sampling simulation experiments, the fitting curve of the prediction price is compared with the real transaction price derived from the optimized model. As a result, the fitting of the optimized model is better as well as the accuracy is higher.

Prediction of Prices for Used Car by using Regression Models

Abstract: For this research, we conducted a comparative study on performance of regression based on supervised machine learning models. Each model is trained using data of used car market collected from German e-commerce website. As a result, gradient boosted regression trees gives the best performance with mean absolute error (MSE) = 0.28. Followed by random forest regression with MSE = 0.35 and multiple linear regression with MSE = 0.55 respectively.

Prediction car prices using qualify qualitative data and knowledge-based system

Abstract: Car pricing using machine learning has a strong relationship with the process of knowledge acquisition for expert systems. Recently, the primary technique for knowledge acquisition has been the time-consuming process of recommendation, posting for car buying or selling on internet market websites. After discovering the data, we can divide that into two types: structured and unstructured that require knowledge-based analysis. This paper will involve the techniques for extraction of meaning, data inference, and rules for qualitative data. The main purpose of the current research is to explore different data types of car data and the objective is to create an automated technique to predict car prices.

3. IMPLEMENTATION

The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. 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, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and its value in the present day scenario. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at.

Disadvantages:

1. Not accurate
2. Not effective

Proposed System:

To overcome this problem we have developed a model which will be highly effective. Machine learning Algorithms are used because they provide us with continuous value as an output and not a categorized value. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.

Advantages:

1. Highly effective

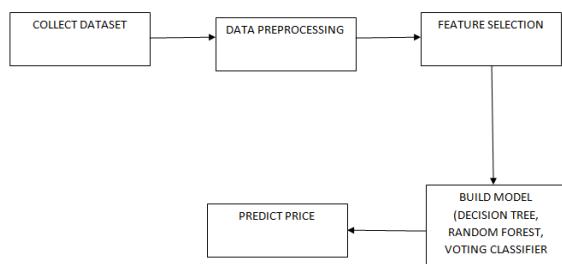


Fig.2: System architecture

To implement this project we have designed following modules

1. Collect dataset
2. Data preprocessing
3. Feature selection
4. Build model (Decision tree, Random forest, Voting classifier)
5. Predict price

4. ALGORITHM

ALGORITHMS:

DECISION TREE:

Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. In other words, we can say that the purity of the node increases with respect to the target variable. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions.

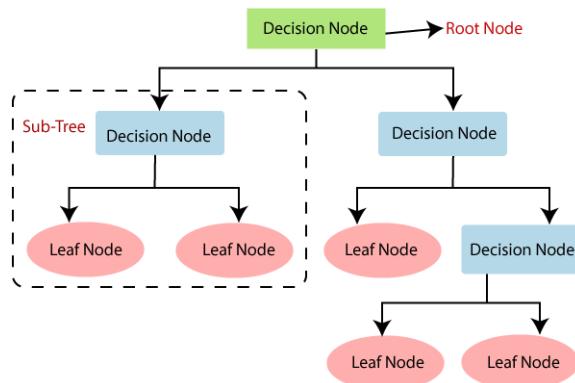


Fig.3: System architecture

RANDOM FOREST:

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. Random Forest is suitable for situations when we have a large dataset, and interpretability is not a major concern. Decision trees are much easier to interpret and understand. Since a random forest combines multiple decision trees, it becomes more difficult to interpret.

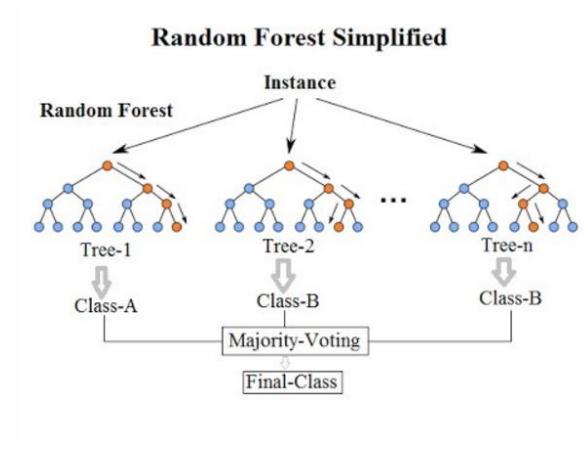


Fig.4: Random forest model

VOTING CLASSIFIER:

A voting classifier is a machine learning estimator that trains various base models or estimators and predicts on the basis of aggregating the findings of each base estimator. The aggregating criteria can be combined decision of voting for each estimator output.

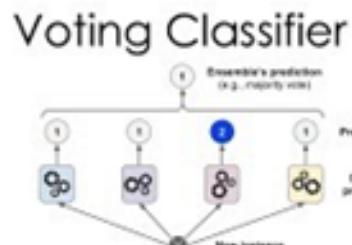


Fig.5: Random forest model

5. EXPERIMENTAL RESULTS

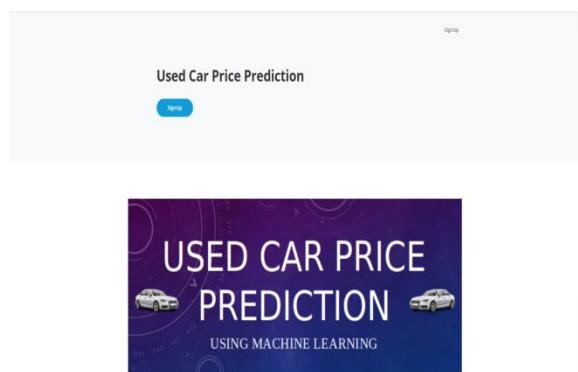


Fig.5: Home screen



Fig.6: User signup

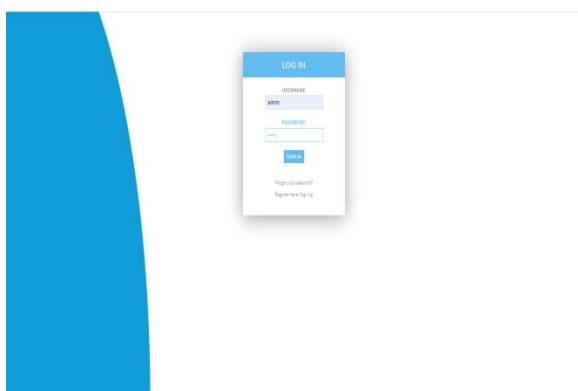


Fig.7: User Signin

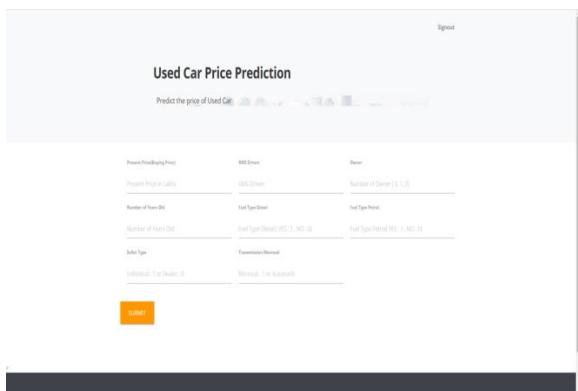


Fig.7: Data uploading screen



Fig.8: Prediction result

6. CONCLUSION

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, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of used car price prediction. This paper compares 3 different algorithms for machine learning : Decision tree, Random forest and Voting classifier.

7. FUTURE SCOPE

In future this machine learning model may bind with various website which can provide real time data for price prediction. Also we may add large historical data of car price which can help to improve accuracy of the machine learning model. We can build an android app as user interface for interacting with user. For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset

REFERENCES

- [1] Sameerchand Pudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques"; (IJICT 2014)
- [2] Enis gagic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, "Car Price Prediction Using Machine Learning"; (TEM Journal 2019)
- [3] Ning sun, Hongxi Bai, Yuxia Geng, Huizhu Shi, "Price Evaluation Model In Second Hand Car System Based On BP Neural Network Theory"; (Hohai University Changzhou, China)
- [4] Nitis Monburinon, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, Pitchayakit Boonpou, "Prediction of Prices for Used Car by using Regression Models" (ICBIR 2018)
- [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)
- [6] Shonda Kuiper (2008) Introduction to Multiple Regression: How Much Is Your Car Worth?, Journal of Statistics Education, 16:3, DOI: 10.1080/10691898.2008.11889579
- [7] Geurts P. (2009) Bias vs Variance Decomposition for Regression and Classification. In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA
- [8] Robert T. (1996) Regression Shrinkage and Selection Via the Lasso. In: Journal of the Royal Statistical Society: Series B (Methodological) Volume 58, Issue 1
- [9] Hastie, Trevor, and Daryl Pregibon. Shrinking trees. AT & T Bell Laboratories, 1990.
- [10] Kim, Tae Kyun. "Understanding one-way ANOVA using conceptual figures." Korean journal of anesthesiology 70.1 (2017): 22.
- [11] Haynes W. (2013) Tukey's Test. In: Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (eds) Encyclopedia of Systems Biology. Springer, New York, NY
- [12] Jaccard, James, Michael A. Becker, and Gregory Wood. "Pairwise multiple comparison procedures: A review." Psychological Bulletin 96.3 (1984): 589.
- [13] Dupac, Václav, ed. Sampling from a finite population. Marcel Dekker, Incorporated, 1981.