Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Prediction of Car Price When Using a Machine Learning Technique

Prof. Pradip Sitaram Ingle[1], Mr. Akshay Shivaji Sonune[2], Mr. Vaibhav Subhash Borade[2]

[1]Assistant Professor, Department of Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

A car price prediction has been high Area of interest research, as it requires noticeable effort and knowledge from the field expert. Considerable Number of different attributes are examined for Reliable and accurate prediction. To build a model for Predict the price of used cars in Bosnia and In Herzegovina we used three machine learning Techniques (Artificial Neural Network, Support Vector Machine and random forest). However, the Mentioned techniques were used to function as one Ensemble. The data used for the prediction were collected from the web portal autopijaca.ba using the web Scrape that was written in PHP programming Language. Respective performances of different Algorithms were then compared to find one that was best Fits available data set. The final prediction model was integrated into the Java application. Furthermore, it is the model was evaluated using test data and accuracy of 87.38% was obtained.

Keywords – Car Prediction, Support Vector Machines, Classification, Machine Learning.

I. INTRODUCTION

This paper Predict the price of used cars in both an important and interesting issue. According to data obtained from the National Transport Authority [1], the number of registered cars between 2003 and 2013 has witnessed a spectacular increase of 234%. With difficult economic conditions, it is likely that sales of used imported (reconditioned) cars and used cars will increase. It is reported in [2] that the sale of Prediction According to information received From the Agency for Statistics of BiH, 921,456Vehicles were registered in 2014, of which 84% ofThey are cars for personal use [1]. This number is increased by 2.7% since 2013, and it is likely that this The trend will continue, and the number of cars will Accurate car prediction involves expert Knowledge, because price usually depends on many Characteristics and factors. Usually most Important are make and model, age,ss Horsepower and mileage. The type of fuel used in Car and fuel consumption per kilometer have a big impact Price of a car due to frequent changes in the price of A fuel. Various features such as exterior color, door Number, type of transmission, dimensions, safety, air Condition, interior, whether it has navigation or not Will also affect the car price. In this article, we Applied various methods and techniques to Achieve higher precision of the price of used car Prediction. Accurate car prediction involves expert Knowledge, because price usually depends on many Characteristics and factors. Usually most Important are make and model, age Horsepower and mileage. The type of fuel used in Car and fuel consumption per kilometer have a big impact Price of

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

a car due to frequent changes in the price of A fuel. Various features such as exterior color, door Number, type of transmission, dimensions, safety, air Condition, interior, whether it has navigation or not Will also affect the car price. In this article, we Applied various methods and techniques to Achieve higher precision of the price of used car Prediction This paper is organized as follows: Section II contains related work in the price field Predictions of used cars. In Part III, the research The methodology for our study is explain. Section IV Elaborates on various machine learning algorithms and Examine their respective perceptions to predict The price of used cars. Finally, in section V, a The end of our work is given, along with Future work schedule.

II. RELATED WORK

Surprisingly, the work with the estimated price of used cars is very recent, but also very sparse. In his MSc thesis [3], Listiani showed that regression mode building using support vector machines (SVM) can estimate the residual price of rental cars with higher accuracy than simple multiple regression or multivariate regression. SVM is Predict the price of used cars using machine learning techniques

Predict the price of used cars has been studied Extensive in various surveys. Listian discussed, In her thesis written for master's thesis [2], thatRegression model built using Support Vector Machines (SVM) can predict the price of a car like Has been leased with better precision than Multivariate regression or some simple multiple Regression. This is on the grounds that Support Vector Machine (SVM) is better at handling datasets With more dimensions and it is less prone to Top fittings and under equipment. The weakness of this Research is that a change of simple regression with More advanced SVM regression was not shown in Basic indicators such as mean, variance or standard

III. DEVIATION

As average deviation or deviation. In another university thesis [4], Richardson worked with the hypothesis that car manufacturers are more willing to produce vehicles that do not weaken quickly. In particular, using a multiple regression analysis, he showed that hybrid cars (cars that use two different power sources to power the car, ie they have both an internal combustion engine and an electric motor) are more able to retain their value than traditional ones. Vehicle. This is probably due to more environmental considerations for the climate and due to higher fuel efficiency. The importance of other factors such as age, mileage, brand and MPG (miles per gallon) were also considered in this study. He collected all his data from various websites.

Table 2. Minimum and Maximum Values

#	CYLINDER VOLUME (CC)	YEAR	PRICE (RS)
Minimum	1000	1988	27, 000
Maximum	2160	2013	825, 000

Wu et al. [5] used neuro-fuzzy knowledge-based system to predict the price of used cars. Only three factors, namely: the car brand, the year it was produced and the engine style were considered in this study. The proposed system gave similar results compared to simple regression methods. Car dealers in the United States sell hundreds of thousands of cars each year through leasing [6]. Selling these cars at the right price has great financial significance for their success. In response to this, the ODAV system (Optimal Distribution of Auction Vehicles) was developed by Du et al. [6]. This system not only estimates the best price for the resale of the cars, but also gives advice on where to sell the car. Since the United States is a large country, the place where the car is sold also has a non-trivial impact on the selling price of used cars. A k-nearest neighbor regression model was used to forecast the price. Since this system was launched in 2003, more than two million vehicles have been distributed through this system

[6]. Gonggi [7] proposed a new model based on artificial neural networks to predict the residual value of private used cars. The main features used in this study were: mileage, manufacturer and estimated service life. The model was optimized to handle non-linear conditions that cannot be done with simple linear regression methods. 3. Methodology Data were collected from << petites announce >> found in daily newspapers such as L'Express [8] and Le In the related work shown above, authors Proposed prediction model based on the single Machine learning algorithm. However, it is noticeable The approach to machine learning algorithms did Do not give remarkable predictions and can be Improved by collecting different machine learning Methods in an ensemble

IV. METHODOLOGY

Data were collected from << petites announce >> found in daily newspapers such as L'Express [8] and Le Defi [9]. We made sure that all the data was collected in less than a month interval, as the time itself could have a noticeable impact on the price of cars. In Mauritius, seasonal patterns are not really an issue, as this does not really affect the purchase or sale of cars. The following data was collected for each car: make, model, cylinder volume (funny, it is considered the same as horsepower in Mauritius), mileage in km, year of manufacture, paint color, manual / automatic and price. Only cars that had their price listed were registered.

Thus, paint color and manual / automatic functions were removed. The data were then further adjusted to remove records where either age (years) or cylinder volume were not available. The model was also removed, as it would have been extremely difficult to get enough records for all the different car models that exist. Although the mileage data were sparse, they were retained as they are considered a

A selection of collected data is shown below in the table

Table 1. Sample Data Collection

#	MAKE	CYLINDER VOLUME (CC)	YEAR	MILEAGE/KM	PRICE (RS)
1	TOYOTA	1300	2007	38000	410000
2	NISSAN	1500	2007	50000	325000
3	HONDA	1500	2005	59000	385000
4	TOYOTA	1000	2007	59000	360000
5	TOYOTA	1300	1989	62665	50000
6	TOYOTA	1500	2008	67000	615000
7	TOYOTA	1500	2008	69000	575000
8	TOYOTA	1490	2006	73000	450000
9	TOYOTA	1600	2006	82000	550000
10	TOYOTA	1000	2006	85000	325000
11	TOYOTA	1500	2000	113000	325000
12	TOYOTA	1500	2000	129000	218000
13	NISSAN	1500	2001	145000	195000

Advertisement

Initially, 400+ records were collected. After further pruning, for example, we kept only the three of the most popular brands in Mauritius, ie Toyota, Nissan and Honda. In particular, we removed all brands for which there were less than ten records. In terms of cylinder volume, for some cars it was given in a range. We then selected the average value for the site. The values are then pretreated in a form suitable for further processing using machine learning techniques. The minimum and maximum values for some numeric functions are shown in Table 2.

The collected raw data set contains 1105 samples. SinceSince data is collected using a web scraper, it is. Many examples that have only a few attributes. In order To clean these samples, PHP script reader Deleted data from the database, perform cleaning and Saves the cleaned samples in the CSV file. CSV The file is later used to load data into WEKA software To build machine learning models [10]. After the cleanup process, the dataset has been Reduced to 797 samples. Especially all brands like Have less than 10 samples and where the price isHigher than 60,000 BAM was removed due to Skew class problem.

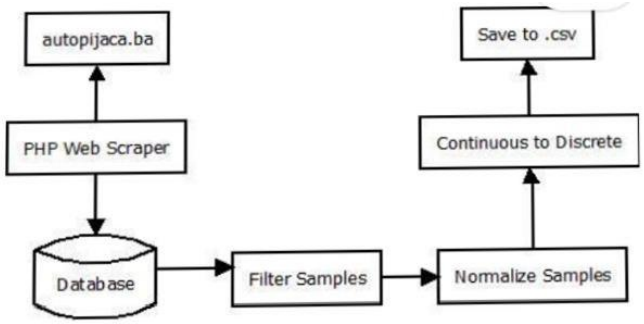

Figure 2. Data gathering and transformation workflow diagram

entire data creation process is displayed in Figure 2.

V. CONCLUSION

Predicting car prices can be a challenging task To the high number of attributes to be Considered for accurate prediction. The big step In the divination process is gathering and Pre-processing of the data. In this survey, PHP Scripts were built to normalize, standardize and clean Data to avoid unnecessary noise for machine learning Algorithms. Data cleaning is one of the processes that Increases prediction performance, but still insufficient for The cases of complex datasets like the one in this Surveys. Use of algorithm for one machine on The accuracy of the dataset was less than 50%. Therefore Ensemble of several machine learning algorithms Has been proposed and this combination of ML Methods get an accuracy of 92.38%. This is Significant improvement over a single machine Learning method approach. However, the downside Of the proposed system is that it uses a lot More computational resources than one machine Learning algorithm. Although this system has achieved amazing results Performance in the car price prediction problem our goal For future research is to test this system to work Successful with different datasets

VI. 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 car rental Application (doctoral dissertation, master's thesis,TU Hamburg-Harburg).

[3]. Richardson, M. S. (2009). Determinants of used car Resale value. Retrieved from: https://digitalcc.coloradocollege.edu/islandora/object/ coccc% 3A1346 [opened: August 1, 2018.]

[4]. Wu, J. D., Hsu, C. C., and Chen, H. C. (2009). An Expert price warning system for used cars Using adaptive neuro-fuzzy inference. Expert Systems with Applications, 36 (4), 7809-7817.

[5]. Du, J., Xie, L., & Schroeder, S. (2009). Practice Price paper – PIN optimal distribution of auction Vehicles System: Use of price warning, Estimation of elasticity and genetic algorithms for Distribution of used vehicles. Marketing Science, 28 (4), 637-644.

[6]. Gongqi, S., Yansong, W., and Qiang, Z. (2011,January). New model for prediction values Of the used car based on the BP Neural Network and Non-linear curve fitting. In measurement technology and Mechatronics Automation (ICMTMA), 2011 third International Conference on (Vol. 2, pp. 682-685). IEEE.

[7]. Pudaruth, S. (2014). Predict the price of used cars Using machine learning techniques. Int. J. Inf. Calculate. Technol, 4 (7), 753-764.

[8]. Noor, K., & Jan, S. (2017). Prediction for vehicles System using machine learning Techniques. International Journal of Computer Applications, 167 (9), 27-31.

[9]. BiH Auto Pijaca. (n.d.), Retrieved From: https://www.autopijaca.ba. [opened August 10, 2018].

[10]. Week 3 – Data Mining with Open Source Machine Learning software in Java. (n.d.),

Retrieved From: https://www.cs.waikato.ac.nz/ml/weka/.[4. August 2018].

[11]. Ho, T. K. (1995, August). Random decision forests. In Document Analysis and Recognition, 1995., Negotiations for the third international conference On (Vol. 1, pp. 278-282). IEEE.

[12]. Russell, S. (2015). Artificial intelligence: A modern Approach (3rd edition). PE.

[13]. Ben-Hur, A., Horn, D., Siegelmann, H. T. and Vapnik, V. (2001). Support vector clustering. Journal of Machine learning research, 2 (Dec), 125-137.

[14]. Aizerman, M. A. (1964). Theoretical basis for The potential function method in pattern recognition Learning. Automation and remote control, 25, 821-837.