International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

Prediction of Used Car Prices Using Artificial Neural Networks and Machine Learning

[1]N Sudha Laxmaiah, [2]Konda Shireesha, [3]Bandaru Prathima

[*1] Assistant Professor, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

[*2,3] Students, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

ARTICLEINFO

Article History:

Accepted: 20 June 2023

Published: 07 July 2023

Publication Issue

Volume 9, Issue 4

July-August-2023

Page Number

47-51

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.

Keywords: Component, Formatting, Style, Styling, Insert

I. INTRODUCTION

Predicting the price of used cars in both an important and interesting problem. According to data obtained from the National Transport Authority [1], the number of cars registered between 2003 and 2013 has witnessed a spectacular increase of 234%. From 68, 524 cars registered in 2003, this number has now reached 160, 701. With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. It is reported in [2] that the sales of new cars has registered a decrease of 8% in 2013. In many developed countries, it is common to lease a car rather than buying it outright. A lease is a binding contract between a buyer and a seller (or a third party –usually a bank, insurance firm or other financial institutions) in which the buyer must pay fixed instalments for a pre-defined number of months/years to the seller/financer.

After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, it is of commercial interest to seller/financers to be able to predict the salvage value (residual value) of cars with accuracy. If the residual value is under-estimated by the seller/financer at the beginning, the installments will be higher for the

clients who will certainly then opt for another seller/financer. If the residual value is over-estimated, the installments will be lower for the clients but then the seller/financer may have much difficulty at selling these high-priced used cars at this over-estimated residual value. Thus, we can see that estimating the price of used cars is of very high commercial importance as well. Manufacturers' from Germany made a loss of 1 billion Euros in their USA market because of mis-calculating the residual value of leased cars [3]. Most individuals in Mauritius who buy new cars are also very apprehensive about the resale value of their cars after certain number of years when they will possibly sell it in the used cars market.

Predicting the resale value of a car is not a simple task. It is trite knowledge that the value of used cars depends on a number of factors. The most important ones are usually the age of the car, its make (and model), the origin of the car (the original country of the manufacturer), its mileage (the number of kilometers it has run) and its horsepower. Due to rising fuel prices, fuel economy is also of prime importance. Unfortunately, in practice, most people do not know exactly how much fuel their car consumes for each km driven. Other factors such as the type of fuel it uses, the interior style, the braking system, acceleration, the volume of its cylinders (measured in cc), safety index, its size, number of doors, paint colour, weight of the car, consumer reviews, prestigious awards won by the car manufacturer, its physical state, whether it is a sports car, whether it has cruise control, whether it is automatic or manual transmission, whether it belonged to an individual or a company and other options such as air conditioner, sound system, power steering, cosmic wheels, GPS navigator all may influence the price as well. Some special factors which buyers attach importance in Mauritius is the local of previous owners, whether the car had been involved in serious accidents and whether it is a lady-driven car. The look and feel of the car certainly contributes a lot to the price. As we can see, the price depends on a large number of factors. Unfortunately, information about all these factors are not always available and the buyer must make the decision to purchase at a certain price based on few factors only. In this work, we have considered only a small subset of the factors mentioned above.

II. RELATED WORK

Surprisingly, work on estimated the price of used cars is very recent but also very sparse. In her MSc thesis [3], Listiani showed that the regression mode build using support vector machines (SVM) can estimate the residual price of leased cars with higher accuracy than simple multiple regression or multivariate regression. SVM is better able to deal with very high dimensional data (number of features used to predict the price) and can avoid both over-fitting and underfitting. In particular, she used a genetic algorithm to find the optimal parameters for SVM in less time. The only drawback of this study is that the improvement of SVM regression over simple regression was not expressed in simple measures like mean deviation or variance. In another university thesis [4], Richardson working on the hypothesis that car manufacturers are more willing to produce vehicles which do not depreciate rapidly. In particular, by using a multiple regression analysis, he showed that hybrid cars (cars which use two different power sources to propel the car, i.e. they have both an internal combustion engine and an electric motor) are more able to keep their value than traditional vehicles. This is likely due to more environmental concerns about the climate and because of its higher fuel efficiency. The importance of other factors like age, mileage, make and MPG (miles per gallon) were also considered in this study. He collected all his data from various websites.

Wu et al. [5] used neuro-fuzzy knowledge based system to predict the price of used cars. Only three factors namely: the make of the car, the year in which it was manufactured and the engine style were considered in this study. The proposed system produced similar results as compared to simple

regression methods. Car dealers in USA sell hundreds of thousands of cars every year through leasing [6]. Most of these cars are returned at the end of the leasing period and must be resold. Selling these cars at the right price have major economic connotation for their success. In response to this, the ODAV (Optimal Distribution of Auction Vehicles) system was developed by Du et al. [6]. This system not only estimates a best price for reselling the cars but also provides advice on where to sell the car. Since the United States is a huge country, the location where the car is sold also has a non-trivial impact on the selling price of used cars. A k-nearest neighbour regression model was used for forecasting the price. Since this system was started in 2003, more than two million vehicles have been distributed via this system [6].

Gonggi [7] proposed a new model based on artificial neural networks to forecast the residual value of private used cars. The main features used in this study were: mileage, manufacturer and estimate useful life. The model was optimised to handle nonlinear relationships which cannot be done with simple linear regression methods. It was found that this model was reasonably accurate in predicting the residual value of used cars.

III. PROPOSED SYSTEM

In this section, we'll go over the many algorithms and datasets that were used to create this module. The model will be trained using a dataset with 92386 records. The value of an automobile is determined by factors such as kilometers travelled, year of registration, fuel type, car model, financial power, car brand, and gear type. We implemented five algorithms because this is a regression problem: Lasso Regression, Ridge Regression, Linear Regression.

A. Lasso Regression

The lasso regression allows you to shrink or regularize these coefficients to avoid overfitting and make them work better on different datasets. This type of regression is used when the dataset shows high multi collinearity or when you want to automate variable elimination and feature selection.

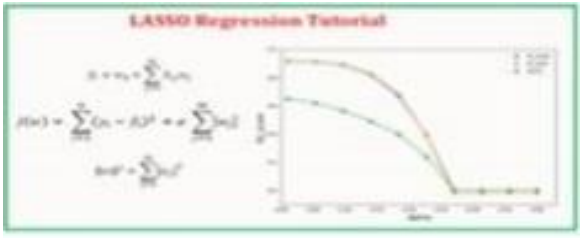

Fig.1.Graphical Representation of Lasso Regression

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}|\beta_j|^{\square}$$

Fig.2. Mathematical Formula of Lasso Regression

Where,

Xij = Features of Y or Independent Variable

Yi = Dependent Variable

βi = Weights or Magnitude shows importance of a feature

λ = minimize the cross-validation prediction error rate

B. Ridge Regression

Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where linearly independent variables are highly correlated. It has been used in many fields including econometrics, chemistry, and engineering. Ridge regression is a sort of linear regression that introduces a little degree of bias in order to improve long-term predictions. Ridge regression is a model regularization technique that reduces the model's complexity.L2 regularization is another name for it. The cost function is changed in this method by including a penalty term.

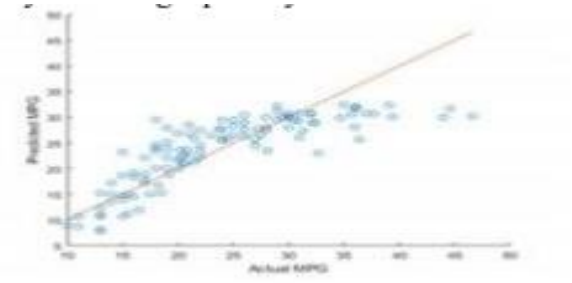

Fig.3.Graphical Representation of Ridge Regression

Ridge Regression penalty is the degree of bias introduced into the model. We may determine it by multiplying the squared weight of each individual label by the lambda.

C. Linear Regression

Quick to train and test as a baseline algorithm.

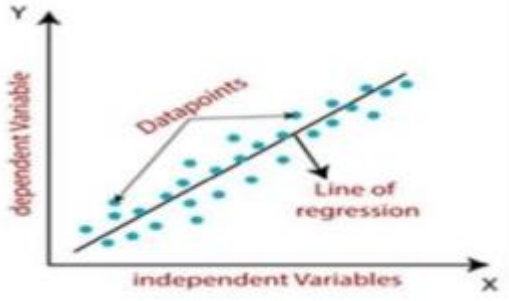

Fig.4.Graphical Representation of Linear Regression

IV. RESULTS AND DISCUSSION

A. Dataset into Data frame

Dataset is given in columns and classifies as

1) Company
2) Model
3) Fuel type
4) Kilometers
5) Year of purchase

	Model	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing XO eRLX Euro III	Hyundai	2007	80,000	45,000 kms	Petrol
1	Mahindra Jeep CL550 MDI	Mahindra	2006	4,25,000	40 kms	Diesel
2	Maruti Suzuki Alto 800 Vxi	Maruti	2018	Ask For Price	22,000 kms	Petrol
3	Hyundai Grand i10 Magna 1.2 Kappa VTVT	Hyundai	2014	3,25,000	28,000 kms	Petrol
4	Ford EcoSport Titanium 1.5L TDCi	Ford	2014	5,75,000	36,000 kms	Diesel

Fig.5. Processed data set sample

Fig.6. Heat Map

V. CONCLUSION

Because of the large number of characteristics that must be examined for an effective prediction, car price prediction will be a difficult assignment. The collecting and preparation of data is the most crucial step in the prediction process. Car data collected from kaggle.com is transformed into CSV format and used to create machine learning algorithms during the research.

In this study, three algorithms were used: Linear, Lasso, and Ridge Regression.SVM classifier separated the data into two portions for training and testing purposes (Support Vector Machine). i.e., 75% of the data was used for machine learning training and 25% of the data was used for machine learning testing. The three machine learning models' accuracy was tested and compared against one another. This is an important comparison between single and multiple groups of machine learning algorithms. As a result, this model will assist in predicting the car's actual price.

VI. REFERENCES

[1]. NATIONAL TRANSPORT AUTHORITY. 2014. Available from: http://nta.gov.mu/English/Statistics/Pages/Archive.aspx [Accessed 15 January2014].

[2]. MOTORS MEGA. 2014. Available from:http://motors.mega.mu/news/2013/12/17/auto-market-8-decrease-sales-new cars/ [Accessed 17 January 2014].

[3]. LISTIANI, M., 2009. Support Vector Regression Analysis for Price Prediction in a Car Leasing Application. Thesis (MSc). Hamburg University of Technology.

[4]. RICHARDSON, M., 2009. Determinants of Used Car Resale Value. Thesis (BSc). The Colorado College.

[5]. WU, J. D., HSU, C. C. AND CHEN, H. C., 2009. An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications. Vol. 36, Issue 4, pp. 7809-7817.

[6]. DU, J., XIE, L. AND 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, Vol. 28, Issue 4, pp. 637-644.

[7]. GONGGI, S., 2011. New model for residual value prediction of used cars based on BP neural network and non-linear curve fit. In: Proceedings of the 3 rd IEEE International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Vol 2. pp. 682-685, IEEE Computer Society, Washington DC, USA.

[8]. LEXPRESS.MU ONLINE. 2014. Available from: http://www.lexpress.mu/[Accessed 17 January 2014]

[9]. LE DEFI MEDIA GROUP. 2014. Available from: http://www.defimedia.info/[Accessed 17 January 2014]

[10]. GELMAN, A. AND HILL, J., 2006. Data Analysis Using Regression and Multilevel Hierarchical Models. Cambridge University Press, New York, USA.

Cite this article as :

N Sudha Laxmaiah, Konda Shireesha, Bandaru Prathima, "Prediction of Used Car Prices Using Artificial Neural Networks and Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 4, pp.47-51, July-August-2023.