Predicting the Price of Second-hand Cars using Artificial Neural Networks

Saamiyah Peerun, Nushrah Henna Chummun and Sameerchand Pudaruth
University of Mauritius
Reduit, Mauritius

ABSTRACT

The number of cars on Mauritian roads has been rising consistently by 5% during the last decade. In 2014, 173 954 cars were registered at the National Transport Authority. Thus, one Mauritian in every six owns a car, most of which are second hand reconditioned cars and used cars. The aim of this study is to assess whether it is possible to predict the price of second-hand cars using artificial neural networks. Thus, data for 200 cars from different sources was gathered and fed to four different machine learning algorithms. We found that support vector machine regression produced slightly better results than using a neural network or linear regression. However, some of the predicted values are quite far away from the actual prices, especially for higher priced cars. Thus, more investigations with a larger data set are required and more experimentation with different network type and structures is still required in order to obtain better predictions.

KEYWORDS
Car price prediction, neural network, linear regression, support vector regression.

1 INTRODUCTION

According to the data obtained from the National Transport Authority (2014), there has been an increase of 254% in the number of cars from 2003 (68, 524) to 2014 (173, 954), as shown in Figure 1. We can thus infer that the sale of second-hand imported (reconditioned) cars and second-hand used cars has eventually increase given that new cars represent only a very small percentage of the total number of cars sold each year. Most individuals in Mauritius who buy new cars also want to know about the resale value of their cars after some years so that they can sell it in the used car market.

Price prediction of second-hand cars depends on numerous factors. The most important ones are manufacturing year, make, model, mileage, horsepower and country of origin. Some other factors are type and amount of fuel per usage, the type of braking system, its acceleration, the interior style, its physical state, volume of cylinders (measured in cubic centimeters), size of the car, number of doors, weight of the car, consumer reviews, paint colour and type, transmission type, whether it is a sports car, sound system, cosmic wheels, power steering, air conditioner, GPS navigator, safety index etc. In the Mauritian context, there are some special factors that are also usually considered such as who were the previous owners and whether the car has had any serious accidents.

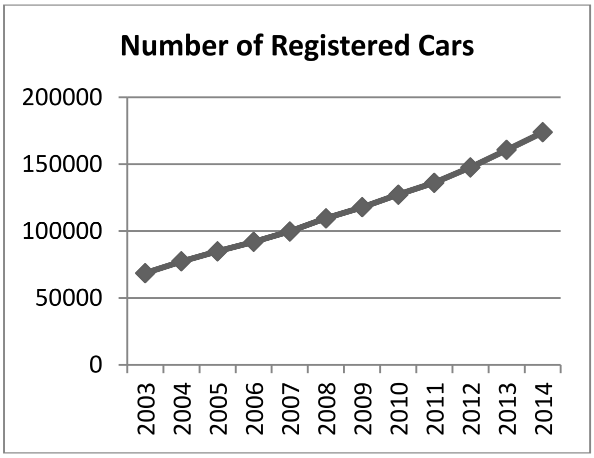

Figure 1. Number of registered cars from 2003-2014

Thus, predicting the price of second-hand cars is a very laudable enterprise. In this paper, we will assess whether neural networks can be used to accurately predict the price of second-hand cars. The results will also be compared with other methods like linear regression and support vector regression.

This paper proceeds as follows. In Section II, various works on neural networks and price prediction have been summarised. The methodology and data collection are described in section III. Section IV presents the results for price prediction of second-hand cars. Finally, we end the paper with a conclusion and some ideas towards future works.

2 RELATED WORKS

Predicting the price of second-hand cards has not received much attention from academia despite its huge importance for the society. Bharambe and Dharmadhikari (2015) used artificial neural networks (ANN) to analyse the stock market and predict market behaviour. They claimed that their proposed approach is more accurate than existing ones by 25%.

Pudaruth (2014) used four different supervised machine learning techniques namely kNN (k-Nearest Neighbour), Naïve Bayes, linear regression and decision trees to predict the price of second-hand cars. The best result was obtained using kNN which had a mean error of 27000 rupees.

Jassbi *et al.* (2011) used two different neural networks and regression methods to predict the thickness of paint coatings on cars. The error for the final thickness of the paint was found to be 2/99 microns for neural networks and 17/86 for regression. Ahangar *et al.* (2010) also compared the use of neural networks with linear regression in order to predict the stock prices of companies in Iran. They also found that neural networks had superior performance both in terms of accuracy and speed compared to linear regression.
Listiani (2009) used support vector machines (SVM) to predict the price of leased cars. They showed that SVM performed better than simple linear regression and multivariate regression. Iseri and Karlik (2009) used neural networks to predict the price of automobiles and achieved a mean square error of 8% compared with 14.4% for regression.

Yeo (2009) used neural networks to predict the retention rate for policy holders of automobile insurance. The neural network was able to predict which customers were likely to renew their policy and which ones would terminate soon. Doganis *et al.* (2006) used artificial neural networks and genetic algorithm in order to predict the sales of fresh milk with an accuracy of 95.4%. Rose (2003) used neural networks to predict the production of cars for different manufacturers.

Thus, we have seen that neural networks have been used successfully for predicting the price of various commodities. Our objective, therefore, in this work, is to use neural networks in a new application, i.e., that of predicting the price of second-hand cars.

3 METHODOLOGY

In order to carry out this study, data have been obtained from different car websites and from the *small adverts* sections found in daily newspapers like L'Express and Le Defi. The data was collected in less than one month interval (i.e. in the month of August in 2014) because like other goods, the price of cars also changes with time. Two hundred records were collected.

The data comprises of different features for second-hand cars such as the year (YEAR) in which it was manufactured, the make (MAKE), engine capacity (ENGINE) measured in cubic centimetres, paint (PAINT) type (normal or metallic), transmission (T/N) type (manual or automatic), mileage (MILEAGE) (number of kilometres the car has been driven) and its price (PRICE) in Mauritian rupees.

Table 1. Snapshot of the car dataset

YEAR	MAKE	ENGINE	PAINT	T/N	MILEAGE	PRICE
2008	Toyota	1400	1	1	70000	315000
2008	Toyota	1400	1	1	70000	315000

2008	Ford	1400	1	0	75000	300000
2005	Mazda	1400	1	1	80000	290000
2009	Citroen	1400	1	0	13000	285000
2005	Mazda	1400	1	1	80000	285000
2006	Renault	1400	0	0	151000	260000
2005	Opel	1400	0	0	125000	230000
2006	Toyota	1400	1	1	75000	230000
2004	Ford	1400	0	0	204000	230000
2004	Toyota	1400	1	0	105000	225000

Table 1 shows eleven records selected from our dataset of 200 records. The range for the year attribute was 2000-2012. A total of fifteen make was studied. Chevrolet and Peugeot had only 3 instances while Toyota has 63 instances. The smallest horsepower in the dataset was 900 and the highest one was 2900. For paint type, 0 stood for normal paint while 1 stood for metallic. For transmission type, a value of 0 means manual transmission while a value of 1 means automatic transmission. The lowest mileage recorded was 2000 km and the highest one was 275,000 km while the price range was 110000 to 685000.

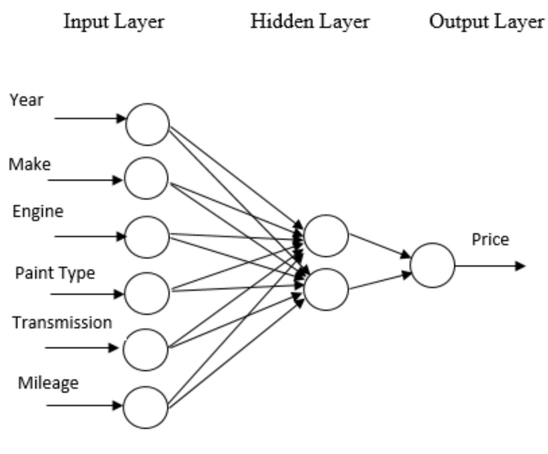

Figure 2. Neural Network Architecture

A neural network with six inputs and one hidden layer containing two nodes was used to predict the price of second-hand cars.

4 EXPERIMENTS AND RESULTS

A large number of experiments have been conducted in order to find the best network structure and the best parameters for the neural network. We found that a neural network with 1 hidden layer and 2 nodes produced the smallest mean absolute error among various neural network structures that were experimented with. However, we found that Support Vector Regression and a multi-layer perceptron with back-propagation produced slightly better predictions than linear regression while the k-Nearest Neighbour algorithm had the worst accuracy among these four approaches. All experiments were performed with a cross-validation value of 10 folds. The results are summarised in Table 2 below.

Table 2. Mean Absolute Errors

Machine Learning Algorithm	Mean Absolute Error (Rupees)
Support Vector Regression	30605
Linear Regression	30828
k-Nearest Neighbour (kNN)	42240
MLP, 500 cycles, learning rate = 0.05	30746

Pudaruth (2014) used only 97 records, 3 make and only 3 features and obtained a mean absolute error of 27000 with kNN (k=1). However, in this work, we have used 200 records with 6 inputs and experimented with more complex approaches. Although, the mean absolute error was slightly higher in our experiments, the value of 30605 can be considered to be a satisfactory outcome as the mean price of the cars was found to be Rs 311586, which is less than 10%. The actual price values obtained from the different sources have been assumed to reflect the true value of the cars but we should point out that these values are estimated by car owners who often do not have much experience in the car business.

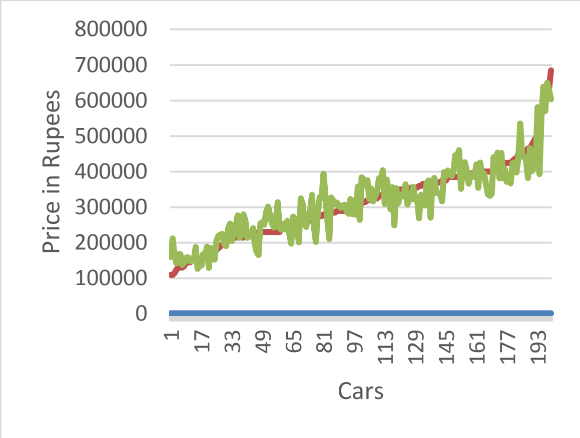

Figure 3. Actual Price v/s Predicted Price using MLP

Figure 3 above shows the variation in actual price (red line) against the variation in the predicted predicted price (green line) using a multi-layer perceptron. The graph shows that as the price gets higher, the deviation from the actual price also increases by a small amount. Nevertheless, the graph also shows that the predictions are also fairly accurate and can relied upon in many cases.

5 CONCLUSION

The aim of this paper was to predict the price of second-hand reconditioned and second-hand used cars in Mauritius. The car market has been increasing steadily by around 5% for the last ten years, showing the high demand for cars by the Mauritian population. There are hundreds of car websites in Mauritius but none of them provide such a facility to predict the price of used cars based on their attributes. Our dataset of 200 records was used with the cross-validation technique with ten folds. The car make, year manufactured, paint type, transmission type, engine capacity and mileage have been used to predict the price of second-hand cars using four different machine learning algorithms. The average residual value was reasonably low for all four approaches. Thus, we conclude that predicting the price of second-hand cars is a very risky enterprise but which is feasible. This system will be very useful to car dealers and car owners who need to assess the value of their cars. In the future, we intend to collect more data and more features and to use a larger variety of machine learning algorithms to do the prediction.

REFERENCES

[1] NATIONAL TRANSPORT AUTHORITY. 2015. Available at: http://nta.govmu.org/English/Statistics/Pages/Archives.aspx. [Accessed 24 April 2015].

[2] Bharambe, M. M. P., and Dharmadhikari, S. C. (2015) "Stock Market Analysis Based on Artificial Neural Network with Big data". *Fourth Post Graduate Conference, 24-25th March 2015, Pune, India.*

[3] Pudaruth, S. (2014) "Predicting the Price of Used Cars using Machine Learning Techniques". *International Journal of Information & Computation Technology,* Vol. 4, No. 7, pp.753-764.

[4] Jassibi, J., Alborzi, M. and Ghoreshi, F. (2011) "Car Paint Thickness Control using Artificial Neural Network and Regression Method". *Journal of Industrial Engineering International*, Vol. 7, No. 14, pp. 1-6, November 2010

[5] Ahangar, R. G., Mahmood and Y., Hassen P.M. (2010) "The Comparison of Methods, Artificial Neural Network with Linear Regression using Specific Variables for Prediction Stock Prices in Tehran Stock Exchange". *International Journal of Computer Science and Information Security*, Vol.7, No. 2, pp. 38-46.

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

[7] Iseri, A. and Karlik, B. (2009) "An Artificial Neural Network Approach on Automobile Pricing". *Expert Systems with Application: ScienceDirect Journal of Informatics*, Vol. 36, pp. 155-2160, March 2009.

[8] Yeo, C. A. (2009) "Neural Networks for Automobile Insurance Pricing". *Encyclopedia of Information Science and Technology, 2nd Edition*, pp. 2794-2800, Australia.

[9] Doganis, P., Alexandridis, A., Patrinos, P. and Sarimveis, H. (2006) "Time Series Sales Forecasting for Short Shelf-life Food Products Based on Artificial Neural Networks and Evolutionary Computing". *Journal of Food Engineering*, Vol. 75, pp. 196–204.

[10] Rose, D. (2003) "Predicting Car Production using a Neural Network Technical Paper- Vetronics (In-house)*"*. Thesis, U.S. Army Tank Automotive Research, Development and Engineering Center (TARDEC).

[11] LEXPRESS.MU ONLINE. 2014. [Online] Available at: http://www.lexpress.mu/ [Accessed 23 September 2014].

[12] LE DEFI MEDIA GROUP. 2014. [Online] Available at: http://www.defimedia.info/ [Accessed 23 September 2014].

[13] He, Q. (1999) "*Neural Network and its Application in IR"*. Thesis (BSc). University of Illinois.

[14] Cheng, B. and Titterington, D. M. (1994). "Neural Networks: A Review from a Statistical Perspective". *Statistical Science*, Vol. 9, pp. 2-54.

[15] Anyaeche, C. O. (2013). "Predicting Performance Measures using Linear Regression and Neural Network: A Comparison". *African Journal of Engineering Research*, Vol. 1, No. 3, pp. 84-89.