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

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.

**EXISTING SYSTEM**

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.

**Disadvantages**

* An existing methodology doesn’t implement DATA PRE-PROCESSING & LABELLING method.
* The system not implemented an effective ML Classifiers for predictions in the datasets.

Proposed System

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.

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 multilayer perception 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.

**Advantages**

* The purpose of linear regression, support vector regression which are more effective for testing and training accuracy.
* In this work, the system will assess whether neural networks can be used to accurately predict the price of secondhand cars.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).