Car Price Prediction Using An Artificial Neural Network

Mehmet Mehdi KARAKOÇ[1*], Gaffari ÇELİK[2] and Asaf VAROL[3]

[1] *Ağrı İbrahim Çeçen University, Vocational School, Department of Computer Technologies, Ağrı, Turkey, mkarakoc04@gmail.com*

[2] *Ağrı İbrahim Çeçen University, Vocational School, Department of Computer Technologies, Ağrı, Turkey, gcelik@agri.edu.tr*

[3] *Maltepe University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, İstanbul, Turkey, varol.asaf@gmail.com*

Abstract

When buying and selling cars, it can be a challenge to assign the correct price. Artificial neural networks, a branch of artificial intelligence, are frequently used for such calculations. In this study we designed two different artificial neural networks for car price forecasting and tested them using data from a car sales website. For data, a software was developed using the C# programming language and the MSSQL Server database management system, also HTMLAgilityPack library was used to read the data on the website. A procedure was written with T-SQL language to digitize the data. As a result of the study, using data from approximately 1000 cars, a success rate of 91.38% was obtained. More data is needed for better results.

Keywords: Artificial Intelligence, Artificial Neural Networks, Car Price Prediction, HTMLAgilityPack.

1. Introduction

Artificial neural networks (ANN) are software or algorithms that solve real life problems by imitating the human thinking and learning system. Artificial Neural Networks applications are mostly used in data classification, prediction, association, filtering and data interpretation processes (Ağyar 2015).

Received:16.08.2020
Accepted:09.12.2020
Published:16.12.2020
[*]Corresponding author: Mehmet Mehdi Karakoç, *Msc*
Ağrı İbrahim Çeçen University, Vocational School, Department of Computer Technologies, Ağrı, Turkey
E-mail: mkarakoc04@gmail.com
Cite this article as: M. M. Karakoç et. al., Car Price Prediction Using An Artificial Neural Network, Eastern Anatolian Journal of Science, Vol. 6, Issue 2, 44-48, 2020.

With the developing technology, ANNs have started to enter every field of our lives day by day. Artificial neural networks are used in almost every field, especially in areas such as image processing, simulation, prediction of social events in the military field; for example; disease diagnosis, treatment methods, preventive medicine, drug development, quarantine processes in the field of health; In the automotive industry, price estimation, production methods, market analysis, test simulators; robotics in electronics, autonomous systems, circuit design; energy production in the field of energy, energy efficiency, energy policies; especially for simulation technologies in space sciences; stock value prediction in banking and finance, product price prediction, economic crisis predictions, management games, exchange rate prediction (Li, Lee, Yao, Lu, & Fan, 2020) (Bourquin Schmidli, Van Hoogevest, & Leuenberger 1998) (Ozsahin & Ozsahin 2020) (Bohr 2020).

Most people have bought or sold a car at some time in their life and the auto trade is a multi-million dollar industry. Therefore, it is essential to determine the right price for both buyer and seller. Prices are usually estimated by looking at those of similar cars advertised on the internet.

In this study, we attempted to estimate car prices using artificial neural networks and set 20 different parameters.

When people buy and sell vehicles, they act by looking at the prices of similar cars. For this purpose one of the most viewed website in Turkey is sahibinden.com. In general, sellers set a price above what they want to sell when giving prices. Seeing this and owning a similar car, another user may think that their vehicle is better than this and set an even higher price. This process causes car prices to rise rapidly over time (Daştan 2016). With the decrease in vehicle production during the COVID-19 process, this high-priced supply has also created its own demand. The

car prices in Turkey is still higher than the inflation rate (Koc & Bilgic 2020) (Ekonomist 2020) (Aymelek 2020).

In the future, it is aimed to develop an artificial intelligence-based price estimation application in order to determine the effect of the mentioned processes on the current prices and to ensure the price balance between the buyer and seller. In the next study, it is planned to investigate the effect of the related sites on the price increase by comparing the prices to be received from the car traders with the prices estimated by the developed program and the prices that the sellers add to the car sales sites (Daştan 2016) (Agarwal & Ratchford 1980).

2. Materials

An artificial neural network (ANN) is an artificial intelligence technique that works by emulating the operation of the human nervous system (Talu & Tatar, 2005). Fig. 1 shows the basic block diagram of an ANN.

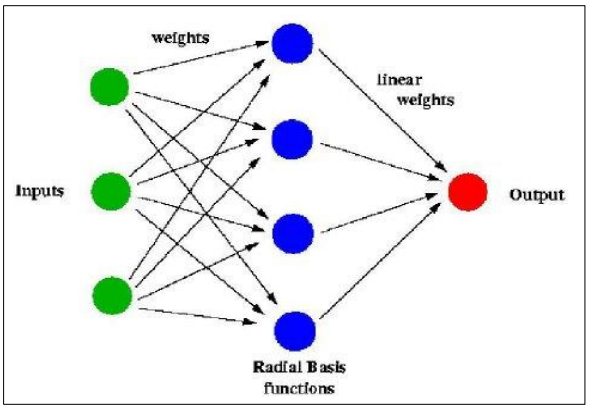

Fig. 1 The basic block diagram of an artificial neural net (Laxmi & Rohil, 2014).

The basic components of an ANN are learning algorithm and activation function. The learning algorithm is the process of finding the desired result by calculating the appropriate weight values with real-life data. The activation function is the component that processes the numerical data coming to neurons to be reduced to a certain interval. ANNs have a layered structure and each layer consists of neurons. An ANN is classified as a single layer network, a multilayer network, a feed forward network where the output of a layer is used as the input of the next layer, and feedback networks where this output is sent as input to the previous layer (Han & Kopacck, 1995) (Bourquin, Schmidli, Van Hoogevest, & Leuenberger, 1998) (Sieniutycz, 2019).

A neural network is made up of nerve cells called neurons. Each neuron has input and output values. Inputs from the external environment or other cells are connected to the cell by means of weights. The net output is calculated by the addition function. The net output is passed through the activation function to calculate the net output. This finds the output value of the cell (Han & Kopacck 1995). Fig. 2 shows a neuron block diagram.

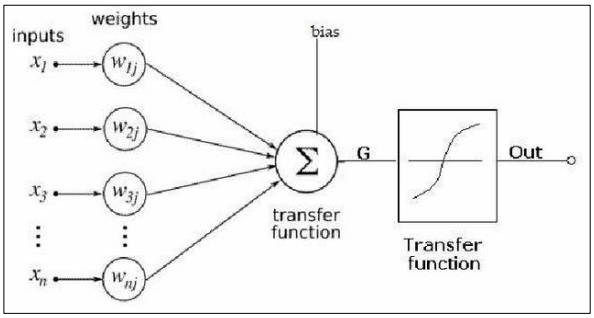

Fig. 2 The basic block diagram of a neuron (Thati, Biswas, & Chowdhury, 2015).

ANNs are used to solve many problems, e.g. expert systems, optimization, classification, pattern recognition and control (Thati, Biswas, & Chowdhury 2015) (Deperlioğlu & Köse 2011) (Elmas 2003).

An ANN usually consists of three layers of neurons; but the number of layers may vary, depending on the problem state (Elmas 2003).

2.1 Input Layer

The input layer is where the input arrives from the outside world to the ANN. In this layer, there are as many cells as the number of inputs, and the inputs are transmitted to the hidden layer without any processing.

2.2 Hidden Layers

Hidden layers transmit the information received from the input layer to the next layer. The number of hidden layers and the number of cells in the hidden layer can vary from network to network. The cell numbers in the hidden layers are independent of the input and output numbers.

2.3 Output Layer

The output layer processes the information from the hidden layer and sends the resulting data to output for the outside world.

The number of cells in the output layer may be greater than one. Each output cell has one output and is connected to all cells in the previous layer.

3. *Methods*

An ANN needs a dataset for training and testing (Elmas 2003). For datasets, we used car advertisements on the most popular car sales web site in Turkey, sahibinden.com (Google Trends n.d.). To access the data, software was developed using the C# programming language and the MSSQL Server database management system. In addition, HTMLAgilityPack was used to read the data on the page. As a result, we obtained data for 1067 cars. This information was first digitized and then normalized to avoid any inconsistencies in the ANN and to balance its weights. Missing and incorrect data were converted into numeric values (Bala & Kumar 2017).

For the purpose of prediction, a multi-layered-feedback ANN was designed in MATLAB containing neurons for 20 inputs (features) and one output (price). The input categories were: brand, series, model, year, fuel, gear, km, body type, motor power, engine displacement, traction, color, warranty, damage condition, plate, from, trade and state. These categories can have a different number of values. For example; body type parameter can has 8 different values as Sedan, Station Wagon, Hatchback with 5 doors, Coupe, Hatchback with 3 doors, MPV, Cabrio, Crossover. Table 1 shows how many values the parameters take. The output data was price.

Table 1 Diversity of input values

Inputs	Diversity of Value
Brand	32
Series	153
Model	477
Year	Didn't replaced
Fuel	3
Gear	3
Km	Didn't replaced
Body type	8
Motor power	107
Engine displacement	95
Traction	4
Color	18
Warranty	2
Damage condition	4
Plate	2
From	2
Trade	2
State	2

Two different ANNs were designed, containing differing numbers of hidden layers and neurons. Some of the data were used for training and some for testing. The first ANN had two secret layers and 30 and 25 neutrons (fig. 3). 70% of the data were used for training and 30% as the test data.

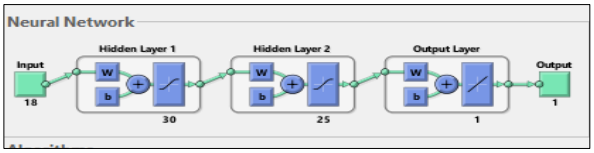

Fig. 3 ANN (Two Hidden Layers)

The second ANN had three secret layers and 30, 15 and 25 neutrons (fig. 4). 60% of the data were used for training and 40% as the test data.

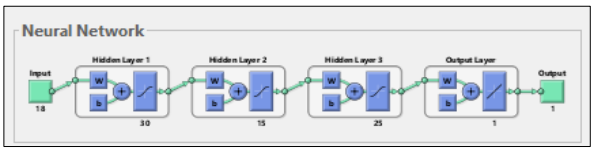

Fig. 4 ANN (Three Hidden Layers).

4. *Resuls*

The success chart of the ANN with two secret layers and 30 and 25 neutrons is illustrated below through a block schema (Fig. 5).

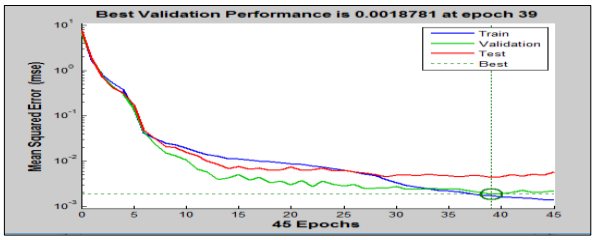

Fig. 5 Success Chart of ANN (Two Hidden Layers).

As shown in fig. 5 above, the program worked for 45 iterations and the minimum squared error of 39 iterations was calculated as 0,0018781. The program was terminated because the error increased in subsequent iterations. Fig. 6 shows the performance of the ANN.

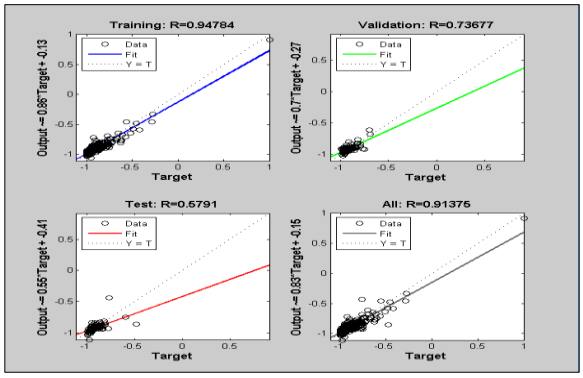

Fig. 6 The performance of the first ANN.

The success chart of the ANN with three secret layers and 30, 15 and 25 neutrons is illustrated below through a block schema (Fig. 7).

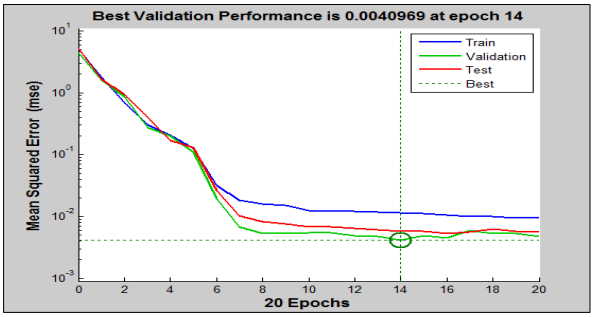

Fig. 7 Success Chart of ANN (three Hidden Layers).

As shown in fig. 7 above, the program worked for 20 iterations and the minimum squared error of 14 iterations was calculated as 0,0040969. The program was terminated because the error increased in subsequent iterations. Fig. 8 shows the performance of the ANN.

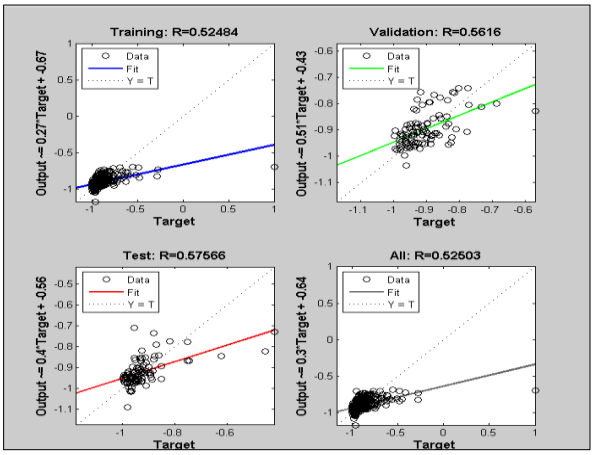

Fig. 8 The performance of the second ANN.

As seen, although the number of layers and neurons increases, when the amount of data used for training decreases, the accuracy also decreases.

In this study, a 91.38% success rate in car price estimation was achieved. In order to achieve more accurate results, a much larger dataset is required.

Since the website blocks more than a certain number of requests for security reasons, more data could not be retrieved. Still, the predictions parallel those of a person. In the next study, it is planned to investigate the effect of the related site on the price increase by comparing the prices to be received from the people who buy and sell cars with the predictions of the developed program and what the sellers add to the website.

5. References

AGARWAL, M. K., & RATCHFORD, B. T. (1980). Estimating Demand Function for Product. *Journal of Consumer Research, 7*(3), 249-262.

AĞYAR, Z. (2015). Use of Artificial Neural Networks (in Turkish). *Engıneer and Machine, 56*(662), 22-23.

AYMELEK, Ü. O. (2020, August 3). *Why have used car prices increased? (in Turkish)*. Retrieved from bbc.com: https://www.bbc.com/turkce/haberler-turkiye-53639568

BALA, R., & KUMAR, D. (2017). Classification Using ANN: A Review. *International Journal of Computational Intelligence Research, 13*(7), 1811-1820.

BOHR, H. (2020). Drug discovery and molecular modeling using artificial intelligence. In A. Bohr, & K. Memarzadeh, *Artificial Intelligence in Healthcare* (pp. 61-83). Academic Press.

BOURQUIN, J., SCHMIDLI, H., VAN HOOGEVEST, P., & LEUENBERGER, H. (1998). Advantages of Artificial Neural Networks (ANNs) as alternative modelling technique for data sets showing non-linear relationships using data from a galenical study on a solid dosage form. *European Journal of Pharmaceutical Sciences, 1*(7), 5-16.

DAŞTAN, H. (2016). Determination of The Factors That Effect Second-Hand Automobile Prices in Turkey by Using Hedonic Pricing Model (in Turkish). *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 1*(18), 303-327.

DEPERLIOĞLU, Ö., & KÖSE, U. (2011). An educational tool for artificial neural networks. *Computers & Electrical Engineering, 37*(3), 392-402.

EKONOMIST. (2020, June 28). *Why did second hand car prices rise? When can prices stabilize? (in Turkish)*. Retrieved from Ekonomist.com: https://www.ekonomist.com.tr/otomotiv/2-el-otomobil-fiyatlari-neden-yukseldi-fiyatlar-ne-zaman-dengelenebilir.html

ELMAS, Ç. (2003). *Artificial Neural Network – theory, architecture, training,*

implementation (In Turkish). Ankara, Turkey: Seçkin Press.

GOOGLE TRENDS. (n.d.). Retrieved November 12, 2018, from https://trends.google.com/trends/explore?q=Arabam.com,Sahibinden.com,Araba.com,arabam.com,tasit.com&geo=TR

HAN, M.-W., & KOPACCK, P. (1995). Optimization Using Artificial Neural Networks. *IFAC Proceedings Volumes, 28*(24), 357-360.

KARCI, A., & DEMIR, M. (2009). Estimation of protein structures by classification of angles between a carbons of amino acids based on artificial neural Networks. *Expert Systems with Applications, 36*(1), 5541-5548.

KOC, C., & BILGIC, T. (2020, February 15). *New-car shortages lead Turkish run on used vehicles*. Retrieved from Autonews.com: https://europe.autonews.com/sales-market/new-car-shortages-lead-turkish-run-used-vehicles

LAXMI, V., & ROHIL, H. (2014). License Plate Recognition System using Back Propagation Neural Network. *International Journal of Computer Applications, 99*(8), 29-37.

LI, B., LEE, Y., YAO, W., LU, Y., & FAN, X. (2020). Development and application of ANN model for property prediction of supercritical kerosene. *Computers and Fluids*(209), 1-18.

OZSAHIN, İ., & OZSAHIN, D. U. (2020). Neural network applications in medicine. In W. Zgallai, *Biomedical Signal Processing and Artificial Intelligence in Healthcare* (pp. 183-206). Academic Press.

SIENIUTYCZ, S. (2019). Complex systems of neural networks. In S. Sieniutycz, *Complexity and Complex Thermo-Economic Systems* (pp. 51-84). Elsevier.

TALU, M. F., & TATAR, Y. (2005). An Edge detection Method for Using Model based Artificial Neural Network (In Turkish). *Journal of Turkey Informatics and Computer Science Foundation, 1*(1).

THATI, A., BISWAS, A., & CHOWDHURY, S. R. (2015). Breath acetone-based non-invasive detection of blood glucose levels. *International Journal on Smart Sensing and Intelligent Systems, 2*(8), 1244-1260.