Science Arena Publications
Specialty Journal of Accounting and Economics
Available online at www.sciarena.com
2016, Vol, 2 (2): 5-14

# NEURAL NETWORKS IN PREDICTING INDIVIDUAL CUSTOMER PROFITABILITY

**Dražena Gašpar[1], Ivica Ćorić[2], Mirela Mabić[3]**
1- Faculty of Economics, University of Mostar, Bosnia and Herzegovina
drazena.gaspar@sve-mo.ba
2- "Hera" d.o.o., Kralja P.Krešimira IV bb, Mostar
ivica.coric@hera.ba
3-Faculty of Economics, University of Mostar, Bosnia and Herzegovina
mirela.mabic@sve-mo.ba

**Abstract:** This paper presents how neural networks can be used for the analysis of current and predicting of future individual customer profitability. Based on historical data stored in the data warehouse, neural networks predict individual customer profitability by defining belonging of particular customer to one of predefined customer segments. Neural networks were used as a tool for analyzing and predicting customer profitability because of their ability to learn from historical data and to make a valid generalization. Since one of the main features of neural networks is their possibility to work with nonlinear and nonfinancial data, it makes them an appropriate tool for predicting individual customer profitability. A presented use of neural networks in predicting individual customer profitability was tested using empirical data from a production company which operates in the Southeast European market.
**Keywords:** Customer Profitability Analysis, Customer Lifetime Value, Forecasting Model, Multi-layered Neural Networks, Neural Networks

## INTRODUCTION

The paper presents and explains use of neural networks in measuring current and future individual customer profitability. The neural networks forecast the activities of individual customers and predict individual customer profitability by defining belonging of particular customer to one of predefined customer segments. Customer profitability analysis (CPA) is a process which identifies the contribution of each particular customer to company's profitability. It does this by measuring contribution of individual customer as the difference between revenues and costs assigned to the particular customer during a given period (Pobrić, 2014, p.190). this is important because customers operate with their specific methodologies causing the allocation of these costs to vary considerably. The main result of this is that each monetary unit of revenue does not participate in profit in the same way. It is the reason why the focus has to be directed at a particular customer, and the measurement of the income (monetary unit "value") which emerges from doing business with this customer. This value is equal to the difference between income and costs resulting from business activities related to the particular customer. But, while revenue presentation is relatively straightforward, costs presentation is a complex problem. Calculation of customer profitability starts by reducing costs, both products and other types of which might strain business transactions, and comparing them to proceeds (Gašpar et al., 2014, pp.425-426). Also, a general framework for defining customer profitability, besides pure financial items, has to include a lot of non-linear and non-financial elements (Gašpar et al., 2012, p. 451). Today, it is obvious that using analysis of customer profitability based only on historical data is not enough. At the global market, the strategic customer management based only on historical data is insufficient. Although analysis based on historical data is valuable and management uses it as the basis for decision making, prospective analysis of customer profitability brings an entirely new knowledge which could be essential for the decision-making process. The individual customer is the base for prospective customer analysis that analysis elements of a business relationship with each customer of the company and makes predictions based on retrospective analysis.

In the last decades, there is a constant increase of ICT support primarily related to storage of an enormous quantity of data in data warehouses and to use of powerful application of neural networks in a business environment. This trend has caused changes in the method of research on customer profitability from a model based on the costs of products or services towards profitability of particular customers.
The neural networks are interconnected assemblies of simple processing elements, units or nodes with functionality loosely based on the biological neuron. The processing ability of the network arose from its inner-unit connection strengths and weights. It is the result of a process of adaptation to or learning from, a set of training patterns (Gurney, 1997, p.13). Neural networks are capable of identifying and absorbing hidden knowledge and patterns of behavior stored in the historical data derived from retrospective customer analysis (Gašpar et.al, 2012, p.452). Neural networks can work equally well with nonlinear and nonfinancial data and that ability makes them a good choice for the predicting of individual customer profitability.
The neural networks are valuable tool in the process of analyzing and predicting individual customer profitability. The reason for this lies in the characteristics of neural network structures and their ability to learn and generalize. Neural networks use historical data in the learning process and making input-output data maps. They can adapt weights if new environmental influences arise, meaning that they are capable of adjusting themselves according to changes in the environment.
The neural networks implementation in predicting individual customer profitability was tested on empirical data from a company which produces and distributes products such as dry fruits, nuts, seeds and cereals for the Southeast European market.

## BACKGROUND

The customer profitability calculation started with a reduction of product costs and proceeds with the recognition and deduction of other types of costs which affect business transactions (van Raaij, 2005; van Raaij et al., 2003; Howell and Soucy, 1990). Continuous evaluation of customer relationships, customer education, negotiation about sales conditions, customers' migration, and termination of business relations with the customer, are only part of activities of the highly intensive process of customer management (Helgesen 2007; Ang and Taylor, 2005; Mittal et al., 2005).
However, previously described approaches to customer profitability belong to retrospective analysis because they use only historical data and analyze past events. Although this type of analysis is valuable, a different approach to profitability based on prospective customer profitability brings a whole new knowledge to the decision-making processes related to customer management. This approach to customer profitability is sometimes called the *lifetime value* (CLV) of the customer (Jain and Singh, 2002) which views customer profitability as a net present value calculation (Gupta et al., 2004). It has become popular due to its forward-looking metrics which incorporate revenues, costs, and customer behavior which together drive future profitability (Kumar and Shah, 2004).
Researchers developed different methods for the calculation of the individual customer value. The aim is ranking of individual clients, segmentation or predicting future values, as can be found in the works of Verhoef and Donkers (2001), Jain and Singh (2002), Stahl et al. (2003), Venkatesan and Kumar (2004), Gupta and Lehmann (2006), Khajvand et al. (2011), Han et al. (2012) and others.
More recently, the academic literature has acknowledged the importance of non-financial effects, this having led to further research investigating the accountability of said effects. It suggests a multidimensional customer profitability measurement rather than one single unified metric (Damm and Rodriguez-Monroy, 2011).
Prospective customer analysis predicts elements of the business activities with the individual customer during his or her tenure as the customer of said company. The method of its predicting is based on retrospective analysis. Neural networks, as one of the machine learning methods, prove their suitability for this type of analysis. Namely, the knowledge stored in historical cases form the basis for neural network learning. It means that neural networks are capable of identifying and absorbing hidden knowledge and patterns of behavior already stored within the historical data of retrospective customer analysis. Neural networks could work equally well with any elements that influence the profitability results, both nonlinear and nonfinancial. Neural networks have proven their capability to describe approximately any continuous function. This characteristic makes them a good choice for the predicting individual customer profitability. The history of modern neural networking started in 1982 when John Hopfield of the California Institute of Technology defined back propagation as a possible method for neural network learning. In 1986, Rumelhart, Hinton and Williams developed the *backpropagation algorithm* (Rumelhart et al. 1986). Similar algorithms were described in the 1974 in doctoral thesis of

Paul J. Werbos from Harvard University, the 1969 book *Applied Optimal Control* by Jr. Arthur E. Bryson and Yu-Chi Hoa, and in the work of Stanford University professor Bernard Widrow and his Ph.D. student, Ted Hoff.
Backpropagation stands for "backward propagation of errors."Although backpropagation is no longer the preferred method for adjusting the weights of a neural network, the method provides insight into how training works by examination of the three steps which it follows: first, marking input nodes, getting a training example, and calculating the output by using the existing weights; second, calculating the error by taking the difference between the calculated result and the expected (actual) result; and third, adjusting the weights and minimizing the error by feeding back errors through the network (Berry and Linoff, 2004, p.229).
The better, and of late more widespread, an algorithm for training neural networks is the *conjugate gradient algorithm* which tests a few different sets of weights and then guesses where the optimum is, using ideas from multidimensional geometry (Berry and Linoff, 2004, p.230).
Over the last decade, there has been a great deal of research related to the application potential of neural networks in customer profitability analysis (Mary and Thangaiah 2012; Berry and Linoff, 2004; Etzion et al.,2005; Gašpar et al., 2012; Gašpar et al., 2014; Penpece and Elma, 2014; Vasant, 2014).

## MATERIALS AND METHODS

Finding the real indicators of individual customer profitability requires a model designed explicitly for this. The most cited model in the literature related to customer profitability is the one developed by Niraj (Niraj et al., 2001). His model is of customer profitability for a wholesale and distribution organization. It is based on activity based costing (ABC) method of cost calculation. The authors adapted the model in a way that instead of ABC used the TDABC (Time –Driven Activity-Based Costing) model. In TDABC model managers directly estimate demands for the resources caused by each transaction, product, or customer.
The aim of a model is to predict correctly one of the three customer layers (segments) for each customer, based on a defined set of input parameters. According to Farris et al. (2006), customers, by type and nature, can belong into three different layers:

1. Top segment customers - the most loyal customers and of the highest value to the organization. The company have to do everything it can to keep those customers. They deserve more attention as a reward for their loyalty.
2. The second segment customers - customers with average to small profits. They have the potential for growth and with additional care they may become top segment customers.
3. The third segment customers - customers the serving of who meets a loss for the company.

The general framework for defining customer profitability, beyond pure financial elements, includes a lot of nonlinear and nonfinancial components which cannot be processed by traditional statistical methods. To ensure a holistic view of customer profitability, the authors propose use of neural networks for predicting individual customer profitability. It requires the existence of an environment that includes a defined framework for the predicting of customer profitability, with the TDABC analysis of costs, and a data warehouse as a data source for the neural network. The TDABC model is used for calculating customer costs. Also, CPA is very sensitive to the quality of data. The implementation of a data warehouse could ensure such quality. The primary goal of a data warehouse is to ensure quality data that are integrated from different sources (internal and external) to support decision making. The ETL (Extraction, Transformation, and Loading) process is a critical stage in data warehouse development that must ensure that data stored in a data warehouse is complete, valid, accurate, consistent, conforming, and integrated.
The authors used a special form of multi-layered neural network. This neural network is called a ***supervised network*** because it requires the desired output to learn. This network creates a model that properly maps the input to the output. In that process, it uses historical data in a way that the model could be used to generate the output in a situation when the desired output is unknown. Hidden layer of variables is the result of nonlinear function activity on a linear combination of input variables. An output layer of neurons is a linear combination of hidden network layers. In the phases of modeling, neural networks are usually represented by an oriented graph. Neural networks expect normalized input and output variables, i.e. all values should be reduced to rank [0,1] or [-1,1].
The aim of developing a neural network model is enabling the process of network learning based on the existing data and verification of a network's ability to predict values of output variables for period t+1, on the basis of data for period t.
The proposed model for predicting customer profitability was tested using empirical data from a company

which produces and distributes products such as dry fruits, nuts, seeds and cereals for the Southeast European market. The data source is a data warehouse generated by the company for the period of 2012 to 2015. All customers are, according to data from Table 1, classified into predefined segments: top customers, customers with average to small profits, and customers who net loss for the company.

*Table 1.Input and output data*

| No. | Data | Description | In/Out variable |
|---|---|---|---|
| 1 | Year | Year to which data refers | N/A |
| 2 | Customer | Customer to which data refers | N/A |
| 3 | Customer size | Customer size (1-small, 2-medium, 3-big) | Input |
| 4 | Customer origin | Customer origin (1-domestic, 2-EU, 3-non-EU) | Input |
| 5 | Costs | Total cost of business with customer in observed year | N/A |
| 6 | Order costs | Total cost of customer's order processing and fulfillment costs in observed year | Input |
| 7 | Shipping costs | Total cost of shipping orders in observed year | Input |
| 8 | Purchase costs | Total cost of purchases and warehousing costs for products delivered to customer in observed year | Input |
| 9 | | Total costs of raising purchase orders to suppliers for delivered products to customer in observed year | Input |
| 10 | | Total costs of receiving shipments from suppliers for delivered products to customer in observed year | Input |
| 11 | Revenue | Total revenue realized by the customer in observed year | Input |
| 12 | No. of deliveries | Total number of deliveries to customer in observed year | Input |
| 13 | No. of products | Total number of different products in customer trade in observed year | Input |
| 14 | No. of delivery places | Total number of different places to which products were delivered for customers in observed year | Input |
| 15 | Net margin | Net margin realized in customer trade in observed year | N/A |
| 16 | No. of returns | Total number of customer goods returned in customer trade in observed year | Input |
| 17 | Value of returns | Total value of returned goods in customer trade in observed year | Input |
| 18 | Value of discount | Total value of the discount given to customers in customer trade in observed year | Input |
| 19 | Customer payment | The average time delay in customer payments in observed year (1- +less than average, 2-more than average) | Input |
| 20 | Customer segment | Determination of segment based on realized customer business results in observed year | Output |

The first step in the implementation of a model (Example 1) should confirm the capacity of classification of used neural network algorithms related to the classification of customers in the adequate segment.
The second step (Example 2) should present the predictive capability of the neural network. The starting point is the same data set but now limited by time. Namely, the idea is to use the same indicators for the prior example but restricted to the first quarter.
The main hypothesis is: *if a network shows the ability to predict correctly the customer belonging to the*

*adequate segment at the end of the year, based on the value of input variables for the first quarter of the year, it means that the network has the predictive capacity.*
The proven hypothesis means that customer management could get a mechanism for better defining a course of action related to the improvement of customer relationship management, ensuring an increase of top segment customers and decreasing the number of customers who lose the company money.
In the both cases, the input set of was divided into:

- Training data (60%),
- Cross Validation data (15%)
- Testing data (25%).

As software tool was used NeuroSolutions version 6.20 for Microsoft Excel of company NeuroDimensions. The tool Express Builder was used for comparing the results of different types of neural network. The number of epochs was 100, where an *epoch* is a complete iteration of the training procedure of neural network.

## RESULTS AND DISCUSSION

Test results of the Example 1 set of data, which were used for testing the classification ability of the model, declared architecture of MLP with the PCA network as the best (Figure 1). A Multi-Layer Perceptron (MLP) is a kind of feed-forward neural network model (i.e. forward directing links), consisting of three layers; the input, hidden, and output layers (Mohamad-Saleh and Hoyle, 2008). This kind of network is called the auto-associative network (Qui et al., 2012). The performance of an MLP depends on its generalization capability, meaning the data it represents. It suggests a need for eliminating correlations in the data before they are being presented to an MLP what can be achieved by applying the Principal Component Analysis (PCA) technique to input data sets before the MLP training process, as well as at the interpretation stage (Mohamad-Saleh and Hoyle, 2008). Figure 2 graphically represents resulting MLP data with a PCA network which consists of 16 nodes in the input layer, one hidden layer with five neurons, and an output layer with one neuron. Learning algorithm is the Levenberg-Marquardt (LM) algorithm as one of the most appropriate higher-order adaptive algorithms for minimizing the MSE of a neural network.
Test results of the Example 2 set of data, which were used for testing the predictive capacity of the neural network, declared the architecture of the Multilayer perceptron (MLP-2-B-L) network as the optimal one (Figure 3). This type of neural network is a ***supervised network*** because it requires the desired output to learn. The aim of this type of network is creation of a model that adequately maps the input to the output using historical data, enabling that the model could be used to produce the output when the desired output is unknown.
Figure 4 graphically represents the resulting Multiplayer Perceptron (MLP) with two hidden layers. The inputs are fed into the input layer and are then multiplied by interconnection weights as they are passed from the input layer to the first hidden layer. Within the first hidden layer, they get summed and processed by a nonlinear function. As the processed data leaves the first hidden layer, is again multiplied by interconnection weights, and then summed and processed by the second hidden layer. Finally, the data is multiplied by interconnection weights once more and processed one last time within the output layer to produce the neural network output.
*Figure 1. Summary of all Networks for Example 1*

*Source: author's calculation*

**Performance Metrics**

| | Training | | | Cross Validation | | | Testing | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model Name** | **MSE** | **r** | **Correct** | **MSE** | **r** | **Correct** | **MSE** | **r** | **Correct** |
| MLP-1-O-M (Multilayer Perceptron) | 0,212518 | 0,421843 | 63,48% | 0,206408 | 0,430879 | 67,54% | 0,20845 | 0,423261 | 64,78% |
| LR-0-B-M (Linear Regression) | 0,185965 | 0,494178 | 68,59% | 0,21151 | 0,411854 | 68,59% | 0,194958 | 0,458382 | 72,01% |
| LR-0-B-L (Linear Regression) | 0,171131 | 0,550502 | 72,12% | 0,217299 | 0,415444 | 73,30% | 0,240005 | 0,393974 | 74,84% |
| MLP-1-B-L (Multilayer Perceptron) | 0,030154 | 0,939425 | 94,63% | 0,045387 | 0,909489 | 92,67% | 0,06371 | 0,870184 | 89,62% |
| PNN-0-N-N (Probabilistic Neural Network) | 0,170736 | 0,554417 | 70,81% | 0,182235 | 0,519332 | 70,16% | 0,161943 | 0,585336 | 73,27% |
| RBF-1-B-L (Radial Basis Function) | 0,094707 | 0,786304 | 83,12% | 0,103456 | 0,769849 | 83,25% | 0,096106 | 0,784134 | 84,91% |
| GFF-1-B-L (Generalized Feedforward) | 0,058126 | 0,877896 | 92,41% | 0,068899 | 0,859287 | 91,62% | 0,081336 | 0,831401 | 87,74% |
| MLPPCA-1-B-L (MLP with PCA) | 0,051559 | 0,892683 | 94,11% | 0,065964 | 0,865181 | 90,05% | 0,055117 | 0,887028 | 92,77% |
| SVM-0-N-N (Classification SVM) | 0,016602 | 0,981925 | 100,00% | 0,131899 | 0,686617 | 76,96% | 0,145606 | 0,6367 | 74,84% |
| TDNN-1-B-L (Time-Delay Network) | 0,05664 | 0,88674 | 92,54% | 0,12405 | 0,752391 | 82,20% | 0,175401 | 0,635397 | 76,10% |
| TLRN-1-B-L (Time-Lag Recurrent Network) | 0,254591 | 0,014381 | 48,04% | 0,259206 | 0,001353 | 49,74% | 0,248583 | 0,013368 | 48,43% |
| RN-1-B-L (Recurrent Network) | 0,210531 | 0,436247 | 66,23% | 0,228729 | 0,373475 | 63,87% | 0,224883 | 0,403956 | 63,21% |
| MLP-2-B-L (Multilayer Perceptron) | 0,035306 | 0,92727 | 94,24% | 0,049492 | 0,89949 | 94,24% | 0,060951 | 0,872182 | 88,68% |
| MLP-1-B-M (Multilayer Perceptron) | 0,167807 | 0,56416 | 71,34% | 0,176171 | 0,542923 | 70,16% | 0,143929 | 0,644128 | 77,99% |
| MLP-2-O-M (Multilayer Perceptron) | 0,1903 | 0,486398 | 69,90% | 0,194035 | 0,469888 | 70,16% | 0,177398 | 0,534208 | 72,64% |
| MLP-2-B-M (Multilayer Perceptron) | 0,184821 | 0,523614 | 66,88% | 0,193779 | 0,484713 | 65,97% | 0,172755 | 0,58537 | 68,55% |
| MLPPCA-1-O-M (MLP with PCA) | 0,203937 | 0,433787 | 65,31% | 0,217985 | 0,375867 | 62,30% | 0,197968 | 0,454802 | 63,21% |
| MLPPCA-1-B-M (MLP with PCA) | 0,167708 | 0,563343 | 71,07% | 0,179825 | 0,531562 | 71,73% | 0,14692 | 0,638685 | 76,73% |
| GFF-1-O-M (Generalized Feedforward) | 0,232279 | 0,523716 | 69,24% | 0,244366 | 0,471951 | 69,11% | 0,195876 | 0,593107 | 75,16% |
| GFF-1-B-M (Generalized Feedforward) | 0,184619 | 0,504677 | 68,59% | 0,196322 | 0,462847 | 68,59% | 0,174551 | 0,545049 | 72,64% |
| RBF-1-O-M (Radial Basis Function) | 0,207207 | 0,406979 | 61,91% | 0,21584 | 0,368851 | 61,26% | 0,19922 | 0,436727 | 62,26% |
| RBF-1-B-M (Radial Basis Function) | 0,233943 | 0,370284 | 58,12% | 0,238768 | 0,321743 | 55,50% | 0,231944 | 0,399222 | 58,18% |
| TDNN-1-O-M (Time-Delay Network) | 0,245103 | 0,242795 | 45,81% | 0,251573 | 0,196924 | 49,74% | 0,238261 | 0,287685 | 47,17% |
| TDNN-1-B-M (Time-Delay Network) | 0,238911 | 0,214562 | 53,66% | 0,238876 | 0,283498 | 51,31% | 0,235159 | 0,244087 | 52,83% |
| RN-1-O-M (Recurrent Network) | 0,174101 | 0,553573 | 69,37% | 0,186547 | 0,516968 | 70,68% | 0,159992 | 0,597347 | 75,79% |
| RN-1-B-M (Recurrent Network) | 0,194143 | 0,461303 | 64,79% | 0,213461 | 0,408537 | 63,35% | 0,200844 | 0,446129 | 68,55% |
| TLRN-1-O-M (Time-Lag Recurrent Network) | 0,201104 | 0,42829 | 65,45% | 0,211308 | 0,390433 | 67,02% | 0,207405 | 0,394308 | 64,78% |
| TLRN-1-B-M (Time-Lag Recurrent Network) | 0,200867 | 0,430883 | 65,05% | 0,221545 | 0,335081 | 62,83% | 0,222166 | 0,311346 | 56,92% |

*Figure 2. Graphical presentation of MLP with PCA neural network*
*Source: author's calculation*

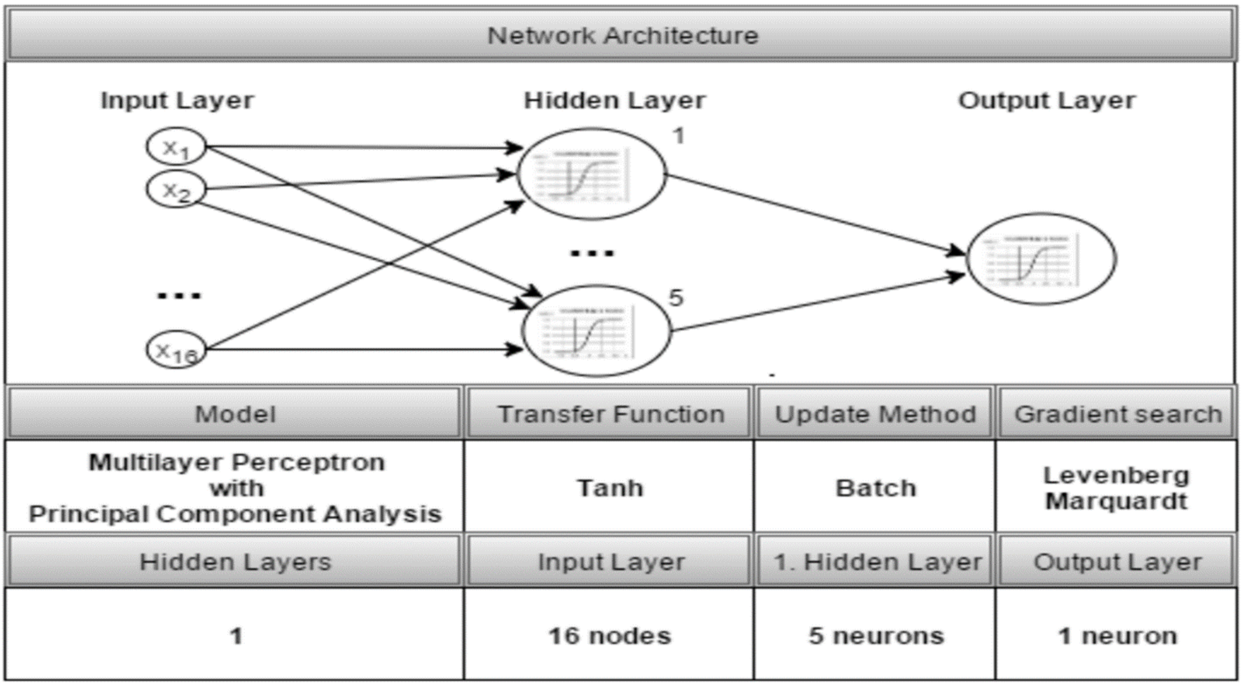


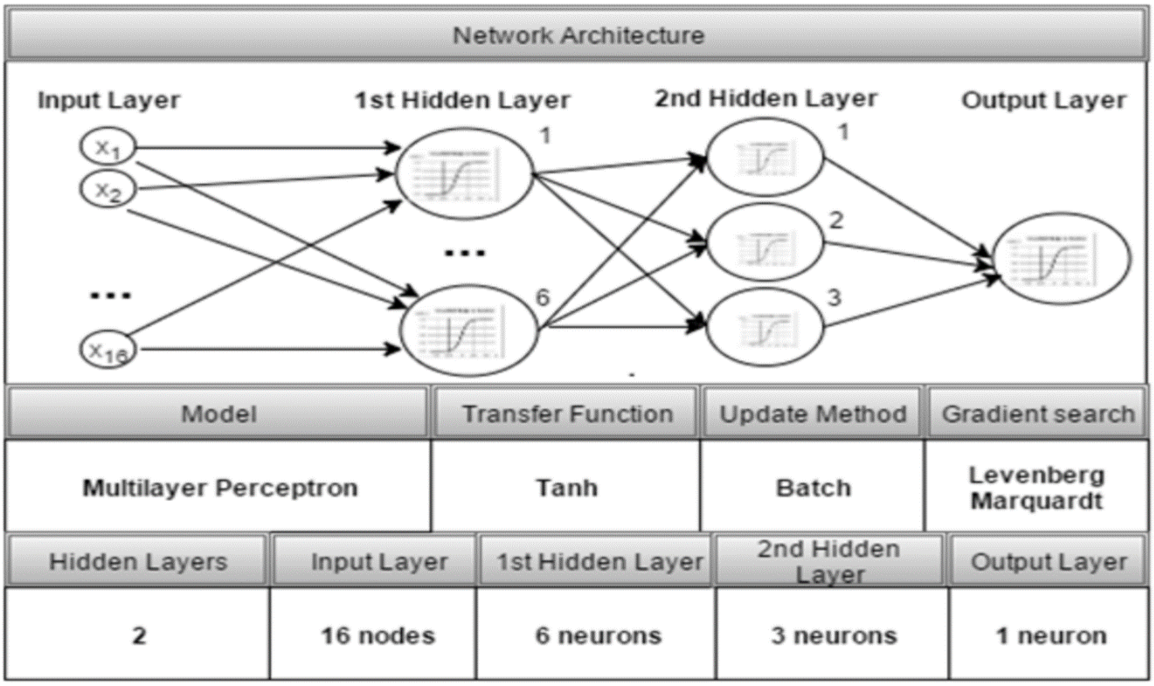

*Figure 3. Summary of all Networks for Example 2*

*Source: author's calculation*

**Performance Metrics**

| Model Name | Training MSE | Training r | Training Correct | Cross Validation MSE | Cross Validation r | Cross Validation Correct | Testing MSE | Testing r | Testing Correct |
|---|---|---|---|---|---|---|---|---|---|
| MLP-1-O-M (Multilayer Perceptron) | 0,023119 | 0,799127 | 63,94% | 0,057128 | 0,573651 | 70,00% | 0,028133 | 0,705184 | 65,67% |
| LR-0-B-M (Linear Regression) | 0,022862 | 0,799571 | 71,57% | 0,047134 | 0,651544 | 70,00% | 0,023562 | 0,748247 | 73,00% |
| LR-0-B-L (Linear Regression) | 0,020844 | 0,819105 | 71,71% | 0,042707 | 0,6918 | 71,67% | 0,023998 | 0,770985 | 73,33% |
| MLP-1-B-L (Multilayer Perceptron) | 0,010861 | 0,910538 | 83,08% | 0,034796 | 0,76401 | 81,67% | 0,015551 | 0,844229 | 79,67% |
| PNN-0-N-N (Probabilistic Neural Network) | 0,011984 | 0,900618 | 74,34% | 0,051918 | 0,617653 | 73,89% | 0,025847 | 0,732992 | 72,67% |
| RBF-1-B-L (Radial Basis Function) | 0,024351 | 0,784963 | 76,70% | 0,049845 | 0,628155 | 73,33% | 0,023327 | 0,753377 | 78,00% |
| GFF-1-B-L (Generalized Feedforward) | 0,011352 | 0,906512 | 83,08% | 0,040248 | 0,719988 | 76,67% | 0,017116 | 0,829247 | 77,67% |
| MLPPCA-1-B-L (MLP with PCA) | 0,008211 | 0,934072 | 88,49% | 0,031036 | 0,794184 | 87,78% | 0,011347 | 0,890418 | 87,00% |
| SVM-0-N-N (Classification SVM) | 0,004928 | 0,984079 | 98,20% | 0,044334 | 0,701921 | 83,33% | 0,02771 | 0,746622 | 76,67% |
| TDNN-1-B-L (Time-Delay Network) | 0,037895 | 0,672264 | 83,36% | 0,06188 | 0,552084 | 74,44% | 0,03652 | 0,60455 | 73,33% |
| TLRN-1-B-L (Time-Lag Recurrent Network) | 0,149828 | -0,03006 | 34,44% | 0,133528 | 0,028662 | 38,33% | 0,152676 | -0,02719 | 35,33% |
| RN-1-B-L (Recurrent Network) | 0,085138 | 0,489145 | 54,58% | 0,090969 | 0,537791 | 57,22% | 0,079132 | 0,530189 | 54,67% |
| MLP-2-B-L (Multilayer Perceptron) | 0,008569 | 0,933509 | 91,11% | 0,032247 | 0,788107 | 88,33% | 0,016397 | 0,844143 | 87,67% |
| MLP-1-B-M (Multilayer Perceptron) | 0,019209 | 0,835219 | 73,47% | 0,051245 | 0,615556 | 73,33% | 0,022994 | 0,756073 | 74,33% |
| MLP-2-O-M (Multilayer Perceptron) | 0,022867 | 0,803237 | 69,17% | 0,051516 | 0,61296 | 71,11% | 0,029337 | 0,693199 | 70,00% |
| MLP-2-B-M (Multilayer Perceptron) | 0,024411 | 0,804077 | 70,00% | 0,048959 | 0,63836 | 67,22% | 0,025159 | 0,743211 | 68,00% |
| MLPPCA-1-O-M (MLP with PCA) | 0,01534 | 0,874216 | 69,58% | 0,0512 | 0,631242 | 66,67% | 0,02768 | 0,726271 | 70,33% |
| MLPPCA-1-B-M (MLP with PCA) | 0,024093 | 0,790193 | 71,39% | 0,050993 | 0,624951 | 70,56% | 0,027743 | 0,700661 | 73,33% |
| GFF-1-O-M (Generalized Feedforward) | 0,02368 | 0,803641 | 61,53% | 0,053876 | 0,605502 | 54,44% | 0,027283 | 0,72134 | 63,33% |
| GFF-1-B-M (Generalized Feedforward) | 0,023898 | 0,789857 | 70,28% | 0,049394 | 0,634528 | 72,22% | 0,023587 | 0,750325 | 71,67% |
| RBF-1-O-M (Radial Basis Function) | 0,029395 | 0,75258 | 60,28% | 0,052704 | 0,610062 | 58,89% | 0,029439 | 0,694409 | 61,00% |
| RBF-1-B-M (Radial Basis Function) | 0,062983 | 0,126879 | 51,39% | 0,080873 | 0,140916 | 52,22% | 0,053129 | 0,148042 | 55,00% |
| TDNN-1-O-M (Time-Delay Network) | 0,021465 | 0,814553 | 57,64% | 0,057851 | 0,547668 | 56,11% | 0,027146 | 0,70752 | 57,00% |
| TDNN-1-B-M (Time-Delay Network) | 0,061099 | 0,331923 | 68,06% | 0,080573 | 0,250551 | 63,89% | 0,053185 | 0,092004 | 61,33% |
| RN-1-O-M (Recurrent Network) | 0,024302 | 0,785626 | 59,86% | 0,048133 | 0,642134 | 54,44% | 0,024433 | 0,737511 | 61,00% |
| RN-1-B-M (Recurrent Network) | 0,027291 | 0,755224 | 68,47% | 0,050214 | 0,626578 | 66,67% | 0,028463 | 0,685636 | 68,00% |
| TLRN-1-O-M (Time-Lag Recurrent Network) | 0,017695 | 0,849635 | 69,58% | 0,054996 | 0,582695 | 65,56% | 0,028548 | 0,702008 | 72,33% |
| TLRN-1-B-M (Time-Lag Recurrent Network) | 0,017388 | 0,85456 | 71,53% | 0,051242 | 0,611037 | 65,56% | 0,023481 | 0,750942 | 66,00% |

*Figure 4. Graphical presentation of MLP neural network*
*Source: author's calculation*




The performances of winning networks for Example 1 and Example 2 are given in Tables 2, 3, 4 and 5. It is evident from Table 3 that the MLP with a PCA network (Example 1) had a minimal error in the case of the training set in 100 epochs, while in the case of the validation set, the network reached the best result in 55 epochs. Table 2 presents the metrics for MLP about PCA. It is obvious that the percentage of correct classifications was 94.11% for the training set, 90.05% for cross-validation, and 92.77% for the testing set. Table 5 shows that the MLP network (Application 2) had a minimal error in the case of the training set in 100 epochs, while in the case of the validation set, the network reached the best results in 99 epochs. Table 4 presents the metrics for MPL where the percentage of correct forecasting was 90.98%for the training set, 88.33% for cross-validation, and 87.67% for the testing set.

*Table 2. Metrics of the Best-Performing networks for Example 1*

| | Training | Cross Val. | Testing |
|---|---|---|---|
| # of Rows | 764 | 191 | 318 |
| MSE | 0.051559 | 0.065964 | 0.055117 |
| Correlation (r) | 0.892683 | 0.865181 | 0.887028 |
| # Correct | 719 | 172 | 295 |
| # Incorrect | 45 | 19 | 23 |
| % Correct | 94.11% | 90.05% | 92.77% |

*Table 3.Final and Minimum MSE for Example 1*

| Best Networks | Training | Cross Validation |
|---|---|---|
| Epoch # | 100 | 55 |
| Minimum MSE | 0.052359567 | 0.094971955 |
| Final MSE | 0.052359567 | 0.103917893 |

*Table 4.Metrics of Best-Performing networks for Example 2*

| | Training | Cross Val. | Testing |
|---|---|---|---|
| # of Rows | 721 | 180 | 300 |
| MSE | 0.008569 | 0.032247 | 0.016397 |
| Correlation (r) | 0.933509 | 0.788107 | 0.844143 |
| # Correct | 656 | 159 | 263 |
| # Incorrect | 64 | 21 | 37 |
| % Correct | 90.98% | 88.33% | 87.67% |

*Table 5.Final and Minimum MSE for Example 2*

| Best Networks | Training | Cross Validation |
|---|---|---|
| Epoch # | 100 | 99 |
| Minimum MSE | 0.077722816 | 0.106239018 |
| Final MSE | 0.077722816 | 0.108770968 |

The presented results of the implementation of neural networks in predicting individual customer profitability, obtained through the use of empirical data from a company which produces and distributes products like dry fruits, nuts, seeds, and cereals for the Southeast European market, proves the capability of neural networks to forecast accurately. Although the percentage of correct forecasting is relatively high, it is obvious that there is room for further improvement. Namely, there are still around 10% of customers assigned to the wrong segment, which could cause unnecessary cost, inadequate customer relationship management, and lower customer profitability.

## CONCLUSION

The paper demonstrates that at the heart of the modern customer profitability analysis is the measurement of profitability on the individual customer level. Only such individual measurement of customer profitability can give a quality base for a detailed analysis of customer profitability distribution inside a company.

The authors showed that neural networks have the potential for analyzing and predicting customer profitability. Their main advantage lies in the fact that neural networks are, by using historical data, capable of learning and adapting their weights if new influences from the environment arise, e.g. it means that they are capable of adjusting themselves according to the changes in the environment.

To ensure that a proposed neural network for forecasting customer profitability will be successful, it is necessary to develop an adequate framework based on the TDBCA method of cost calculation and to ensure the quality of the data source by implementing a data warehouse.

The aim of future research should be testing the importance of input variables and finding possibilities for extending the output set of variables. Further research should concern an analysis of hybrid models i.e. integrating neural networks and genetic algorithm to improve the accuracy of the predicting. The authors are aware of the necessity to test the model on a greater number of similar companies.

The overall conclusion is that the use of neural networks in the predicting of individual customer profitability proposed in this paper is a good starting point for future research in that area.

## REFERENCES

1. Ang, L.& Taylor, B. (2005). Managing customer profitability using portfolio matrices. *Journal of Database Marketing & Customer Strategy Management,12(5)*, 298-304.

2. Berry, M.J.A, & Linoff,G.S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*, 2nd edition, Wiley Publishing, Inc., Indianapolis, Indiana, USA.

3. Damm, R.,& Rodriguez-Monroy, C. (2011). A review of the customer lifetime value as a customer profitability measure in the context of customer relationship management. *Intangible Capital*, 7(2): 261-279.

4. Etzion, O., Fisher, A., & Wasserkrug, S. (2005). E-CLV: A modeling approach for customer lifetime evaluation in e-commerce domains, with an application and case study for online auction. *Information Systems Frontiers, 7*, 421–434.

5. Farris, P.W, Bendle, N.T, Pfeifer, P.E, & Reibstein, D.J. (2006). *Marketing metrics*, Wharton School Publishing, USA.

6. Gašpar, D., Ćorić, I., & Mabić, M. (2014). Data Mining in Customer Profitability Analysis. In Proceedings of 7th International Conference for Entrepreneurship, Innovation and Regional Development (ICEIRD), Cyprus, Greece.

7. Gašpar, D., Markić, B. & Ćorić, I. (2012). Machine Learning in Customer Profitability Forecasting. In *Proceedings of 16th International Research/Expert Conference "Trends in the Development of Machinery and Associated Technology"* TMT 2012, Dubai, UAE.

8. Gupta, S., Lehmann, D.R., & Stuart, J.A. (2004). Valuing customers, *Journal of marketing research 41 (1)*, 7-18.

9. Gupta,S., & Lehmann, D. (2006). *Managing customers as investments*. Upper Saddle River, NJ:

Wharton School Publishing.

10. Han,S.H.,Lu,S.X.,& Leung,S.C.H. (2012). Segmentation of telecom customers based on customer value by decision tree model. *Expert Systems with Applications, 39*, 3964–3973.

11. Helgesen, Ø. (2007). Drivers of customer satisfaction in business-to-business relationships: A case study of Norwegian fish exporting companies operating globally. *British Food Journal, 109(10)*, 819-837.

12. Howell R.A., & Soucy S.R.. (1990). Customer profitability; as critical as product profitability. *Managment Accounting*, October, 43-47.

13. Jain,D., & Singh,S.S. (2002).Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing, 16(2)*, 34–46.

14. Khajvand, M., Zolfaghar,K., Ashoori,S.,& Alizadeh,S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behaviour: Case study. *Procedia Computer Science, 3*, 57–63.

15. Kumar, V.,& Shah,D. (2004). Building and sustaining profitable customer loyalty for the 21st century, *Journal of retailing 80(4)*, 317-329

16. Mittal, V., Anderson, E.W., Sayrak, A., & Tadikamalla, P. (2005). Dual Emphasis and the Long-Term Financial Impact of Customer Satisfaction. *Marketing Science*, 24(4), 544-555.

17. Mohamad-Saleh, J., & Hoyle, B.S. (2008). Improved Neural Network Performance Using Principal Component Analysis on Matlab. *International Journal of the Computer, the Internet and Management* 16(2), 1-8.

18. Niraj, R., Gupta, M., & Narasimhan, C. (2001). Customer profitability in a supply chain. *Journal of Marketing, 65*, 1–16.
19. Penpece,D.,&Elma,O.E. (2014). Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry: A Case Study in Turkey. International Journal of Trade, Economics and Finance, 5( 5), 44-51.
20. Pobrić, A. (2014). Measuring customer profitability: The applicability of different concepts in practice. *Ekonomika preduzeća, 62( 3-4)*, 187-200.
21. Rumelhart, D.E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation, *In: Book Parallel distributed processing: explorations in the microstructure of cognition*, 1, 318-362, MIT Press Cambridge, MA, USA.
22. Stahl,H.K., Matzler,K.,& Hinterhuber,H.H. (2003). Linking customer life-time value with Shareholder value. *Industrial Marketing Management, 32*, 267–279.
23. Van Raaij, E.M. (2005). The strategic value of customer profitability analysis. *Marketing Intelligence and Planning, 23* (4), 372-381.
24. Van Raaij, E.M., Vernooij, M. and Van Triest, S. (2003). The implementation of customer profitability analysis: A case study. *Industrial Marketing Management, 32(7)*, 573-583.
25. Vasant, P.M. (2014). *Handbook of Research on Novel Soft computing Intelligent Algorithms: Theory and Practical Applications*. A volume in the Advances in Computational Intelligence and Robotics, IGI Global, USA.
26. Verhoef,P.C.,& Donkers,B.(2001). Predicting customer potential value an application in the insurance industry. *Decision Support Systems, 32*, 189–199.
27. Venkatesan,R.,& Kumar,V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing, 68*, 106–125.