Customer Segmentation Using Neural Networks

DAJANA ĆORIĆ
Ph.D. Student
Faculty of Economics & Business
University of Zagreb
Trg J. F. Kennedy 6
ZAGREB, CROATIA
dajana.coric@gmail.com

KATARINA ĆURKO
Faculty of Economics & Business
University of Zagreb
Trg J. F. Kennedy 6
ZAGREB, CROATIA
kcurko@efzg.hr

MLADEN VARGA
Faculty of Economics & Business
University of Zagreb
Trg J. F. Kennedy 6
ZAGREB, CROATIA
mvarga@efzg.hr

Abstract: This paper shows how neural networks can be useful in process of predicting customer segments. Customer segmentation provides to companies insight in differentiating profitable from non-profitable customers so that companies can act properly on each group. The aim of this paper is to determine how neural networks are able to predict belonging of each customer to specific segment based on data from data warehouse. This paper presents research on analysing data set of customers of the company which take care of production and distribution of the range of products from a set of nuts, grains and dried fruit. As a result will be three groups of customers: group containing the most profitable customers, group of customers with a profit close to zero, and group of the least profitable customers. Depending on where specific customer belongs, company will use different marketing activities in order to retain them, move them to more profitable segment or ultimately reject them. Customer activity in first quarter will be used to predict segment in which customer belongs at the end of the year. This will show predictive ability of neural networks and how reliable they are in process of customer segmentation using incomplete data.

Key-Words: customer, profitability, neural networks

1 Introduction
The aim of this paper is the use of neural networks in the process of customer segmentation. The interest in neural networks was boosted by the application development at different fields: finance, medicine, geology, physics, and other areas of human work and interests where we can find problems related to anticipation and classification. This application is not coincidental. It is based on key features of neural networks: the power and ease of use. The power of neural networks is reflected primarily in the nonlinearity. Linearity watches systems through the objective function, which are optimized to the limits given in the form of equations and / or inequations and with no negative
decision variables. This way of seeing the world uses relatively simple techniques. Ease and assumptions underlying the linear modeling are not sufficiently adequate for modeling many systems in the real world. The behavior of such systems is characterized by non-linear function with a large number of side variables. Neural networks found an area of their application exactly in this segment of real world.

Its applicability this network thanks to the relative ease of use. Main task of users consist in collecting representative data and hiring internal algorithms of network which recognize the data structure, process them and interpret the results. The user does not have to possess a high level of knowledge to successfully use them. Therefore, the application of that such a powerful mechanism is very wide. They are used in almost all the problems for which there is a relation between input (independent) variables and output that need to be predicted.

The calculation of the profitability of individual customers or groups of customers provides the basis for a deeper analysis of the distribution of customer profitability within the company. Segmentation based on profitability provides the basis for the implementation of management strategies for specific groups of customers.

However, this way of managing customer base is based on a retrospective analysis of data. For the management of company, in terms of constant and dynamic changes in the market, predictive analysis would provide the base for proactive action and manage customers in more effective and efficient way.

The aim of the work is therefore to determine how neural networks are able to predict belonging of each customer to the appropriate segment and on the base of incomplete data from the current operating indicators of the company and with the knowledge gained through the process of learning networks on historical data customer segmentation.

2 Methodology of neural networks
A neural network is a massively parallel distributed machine made up of simple processing units, which has the natural ability of storing experiential knowledge and making it available for use. It is based on imitating the brain as follows:
1. Knowledge is collected from network environment through the learning process.
2. Power of interneuron connections, known as synaptic weights, is used to store adopted knowledge.
3. In artificial neural networks, information is processed in artificial neurons, which are the basic element of a neural network data processing. They consist of three main components which include the weights, threshold and activation function. [1]

An artificial neural network consists of a series of artificial neuron that communicate by sending signals to each other via a large number of connections. Artificial neuron, just as natural, receives certain input parameters via links, each of which has its own weight. Weights wK1, wK2, ..., wKP can have negative values. Each neuron has a threshold of value. The sum of input weight is reduced by the threshold of neuron and forms a so-called trigger signal of neuron. This trigger signal is an input parameter of the activation function of neuron. Activation function produces an output neuron signal.

All neurons can be divided into input (such as neurons that receive signals in the eyes), output (such as eyelid during blinking) and hidden neurons (neurons that perform a number of mid functions mediating between the input and output neuron). Only interconnected they can play the role assigned to them.

Weighting function of network are its key elements and represent indicators of the importance and impact of an individual input parameter on the result of the network. They multiply the value of inputs and so multiplied passed to the function of summarizing. Summarizing is the process that adds all the inputs and as a product gives unique input neuron signal.

The last stage of processing neural network is the transformation in the output signal using a transformation function. The most frequently are used the following four functions: Unit step (threshold), Sigmoid, Piecewise Linear and Gaussian. For the purposes of this study is used a non-linear sigmoid function whose formula is: [5]

$$f(x) = \frac{1}{1 + e^{-ax}}, \quad 0 \leq f(x) \leq 1. \quad (1)$$

The link between input and output elements of the neural network cannot be unambiguously defined. One of the main features is that the network itself
teaches about the nature and meaning of the connections on the basis of training. The power of network lies in its ability to recognize the connection between the input and output variables, and this requires training of network. Training is the process in which the network is learning. Learning can be divided into two categories: supervised and unsupervised. For the purposes of supervised learning prepared sets of historical data is used, so network learns on the basis of system behavior in the past. In the case of unsupervised learning network has no details of the past, but there are only input stimulants of the network. The network learns from a set of input data in a way that recognizes the properties or the correctness of the input data.

2.1 Backpropagation algorithm
The best known and simplest learning algorithm of neural networks is backpropagation algorithm. [4] Backpropagation procedure forms a vector error. Point of vectors are provided along the descending line of the current point, so by moving that line in small steps the error is reduced. The hard part is to determine the length of these sequences. Longer steps can quickly converge local minimum, but they can also "skip" a minimum and make a wrong track. In contrast to larger steps, very small steps can go in the right direction, but this will require a large number of iterations. In practice, size of steps is proportional to the slope and to the variable which is called the learning rate. Backpropagation algorithm progresses through iterations. Through each iteration, the network is charging input variables and output variables are compared with the actual output of the network. Error is used to adjust the weighting factors and the process is repeated. [5]

Therefore, once the input variables are selected, the previously mentioned functions regarding the network design could be systematized in the following steps:
• Selecting an initial configuration, which typically consists of one hidden layer with the number of neurons equal to half the sum of the input and output variables.
• Implementing a certain number of experiments with each configuration while retaining the best network, taking the criterion as mistake of selective data set. It is necessary to conduct a sufficient number of experiments with each network configuration to overcome the configurations that are finding local minimums, and preferably practice of resampling.

• In each experiment, if the network does not show satisfactory performance (under-learning) [6], it is necessary to try adding more neurons in the hidden layer of the network. If this does not help, it is necessary to try to add a new hidden layer.
• If the function of selection error begins to increase (over-learning) it is necessary to try to reconfigure the network by subtracting the individual neurons from the hidden layer or even the entire hidden layer in complex network constructions.
• Once you determine effective network configuration, it is necessary to rearrange the existing data in new sets, and generate a new network starting from the initial configuration through previous training. [5]

3 Previous researches
Successful application of neural networks is present in many areas, eg. health care, military, business, education, and other. Here are just some examples:

- Detection of explosives in baggage at airports,
- Identification of cloud types on the basis of satellite images,
- Signal processing, e.g. for the detection of radar,
- Speech recognition using integrated neural networks,
- Pattern recognition in e.g. sorting mail according to postal codes,
- Text to speech conversion,
- Applications in finance (e.g. for the predictions of the stock markets, portfolio choice, trading on the stock exchange, risk assessment, etc.),
- Application in marketing (e.g. for the customer segmentation, forecasting customer choice) and other parts of the business. [7]

According to research of Wong et al. (1997), the largest share of business applications of neural networks in the last 10 years belongs to production (53.5%), followed by finance (25.4%) and marketing and other areas. Results of application of neural networks show that those in most cases give better results than the traditional method of computing and advantages of using this method can be checked in the reports of many companies that use them, for example, reports of Company Z-Solutions on the
application of neural networks to meet the needs customers in healthcare, and others. [7]

Salchenbergert studied the prediction of failure in savings and the neural network was compared with logistic regression. The neural network had significantly surpassed logistic regression. For example, for a period of 18 months, forecasting logistic regression has achieved success rate of 83.3 to 85.4%, while the neural network achieved a success rate of 91.7% [8]

Coats and Fant had predicted bankruptcy of firms by comparing neural networks and MDA (Multivariate Discriminant analysis). Neural networks have resulted in 95.0% accuracy, while the MDA resulted in an accuracy in the range of 83.7% to 87.9%. [9]

Worldwide famous company Google also uses neural networks. Google has released some new research about its efforts to maximize performance and minimize energy use at data centers through machine learning today. [10]

Young researcher from Google, Jim Gao, published his research in paper called “Machine Learning Applications for Data Center Optimization”. As a conclusion he implied that accelerating growth in DC (Data Centers) complexity and scale is making energy efficiency optimization increasingly important yet difficult to achieve. Using the machine learning framework developed in his paper, he was able to predict DC (Data center) PUE (Power Usage Effectiveness) within 0.004 +/- 0.0055, approximately 0.4% error for a PUE of 1.1. Actual testing on Google DCs indicate that machine learning is an effective method of using existing sensor data to model DC energy efficiency, and can yield significant cost savings. Model applications include DC simulation to evaluate new plant configurations, assessing energy efficiency performance, and identifying optimization opportunities. [11]

4 General framework for customer segmentation

The main activity of the analyzed company is the production and distribution of the range of products from a set of nuts, grains and dried fruit. These products observed company manufactures and distributes to a set of customers who were taken from a database that will be used in the empirical model in obtaining the desired results.

The customer database of the observed company consists of 2 291 customers through the period of four years, 2009 - 2012.

Following information were observed from the customers: turnover of the customer, the number of orders that are made during a given time period, the number of different purchased products, the number of delivery locations, the gross margin, the number of transactions to return goods in the observed period, and the value the return of goods.

4.1 Defining a framework for customer segmentation

The starting point for customer segmentation makes the calculation of the current profitability of individual customer. Customer profitability analysis is a process of three step process that includes:

1. Measurement of contribution or direct customer profit.
2. Defining quantitative measures that reflect individual customer affiliation to a certain group.
3. Evaluation. [12]

The calculation will be made over three data sets from a period of four years. Current profitability, input data for its calculation, and other data that make up the independent variables, will do the data basis for the application of neural networks in an attempt to build the model which will made customer segmentation.

Independent variables of the general model for predicting profitability, input variables, will be classified into the following groups according to their similarities:

- C – Costs,
- F - Features of customers,
- P – Prices.

The output variable will be the following, arbitrarily selected by the authors:

- Y1 - Belonging to a specific segment.

The program that will be used is NeuroSolutions.
4.2 Modeling a neural network for customer segmentation

The following are detailed descriptions of input and output variables.

4.2.1 Input variables

Group of costs, includes variables which at the individual customer express costs incurred in the various stages of the business relationship between individual customer and the company. Of all the costs used will be those relating to cost of sold goods. There are many variables from a set of features of customers. Previously defined costs are just as dependent in part on the features of customers. The size and complexity will be as essential features, and will be shown by the following characteristics arbitrarily selected by authors:

- K1 – Turnover of the customer,
- K2 - The number of orders that the customer made during a given time period,
- K3 - Number of different purchased products,
- K4 - Number of delivery locations,
- K5 - The gross margin achieved in turnover with the customer,
- K6 - Number of transactions of returning the goods in the observed period,
- K7 - Value of returned goods.

Price of the product and its associated gross margin, makes the following important factor of customers’ profitability. Product of product prices and quantity of sales gives revenue. In calculating customer profitability of realized revenue per customer, costs are deducted and is obtained the gross profit. Defined is the following new input variable of the model:

- C1 - Individual discount.

The variable contains the total discount that the customer gets in individual transactions on the basis of previously agreed rate discount in arranging a business relationship. It is common practice in this way of doing business to contract quantity discounts which the customer acquires purchasing above the agreed amount or value of certain products.

4.2.2 Output variable

Output variables are the product of the model for prediction. The model is based on the classification of individual customer in a particular segment. From this standpoint, there is an output variable:

- Y1 - Belonging to a specific segment.

Consideration on costs and their distribution to customers showed that in the customer database is a certain part of profitable and unprofitable customers. The degree of differentiation of these two groups of customers depends on several factors, and graphical presentation of distribution was showed on the best way by Kaplan with so-called "Whale curve" or “Kanthal curve”. Kanthal curve is shown in Fig. 1. [13]

![Fig. 1. Kanthal curve](image)

From the standpoint of Kanthal curve it is relatively easy on the basis of data on realized profits in business with individual customer, divided it into one of the three characteristic segments. The ordinate of the graph expresses the cumulative profit system, expressed as a percentage share of the total profit, while the abscissa shows a summary of the customer base ranked by size of realized profits. [9] The curve gives interesting results: 40% of customers is profitable and brings 250% of the total profit. Last 10% generates the loss of almost 150% of the total profit and the rest of the customer base is around breakeven point, 50% of them (curve ends at a point 100% of the profit). The output variable "Belonging to a specific segment" will have three values that have been explained by Kanthal curve:

- 40% of the most profitable,
- 50% with a profit close to zero,
- 10% of the least profitable. [14]
5 Results
All customer data, 2291, are distributed in three data sets, and they are: training data (data from 2009 and 2010), selection data (data from 2011), and testing data (data from 2012). In total we have 951 training data, 511 selection data, 829 data for testing, which together make up a total of 2291 data used in the work of the neural network. Group of customers by years for particular data set are chosen arbitrarily.

There are three data sets because neural network needs one set on which it will train, another set for Cross Validation of data and, finally, test set on which it will show how good it has or hasn’t learned to segment customers.

What is defined by term data?
Data makes a vector composed of the following elements:
- Year,
- Customer,
- Total cost of goods sold,
- Overall size of transactions with the customer,
- The number of orders for the year,
- Number of different products purchased,
- Number of delivery locations,
- Gross margin achieved in turnover of the customer,
- The number of transactions of return goods in the observed period,
- The value of returned goods, and
- The value of discount for the customer.

The output variable represents belonging to a specific customer segment, and consists of three parts, according to Kanthal curve:
- segmentation (plus) - the most profitable customers
- segmentation (zero) - customers with low positive or negative profitability
- segmentation (minus) - very unprofitable customers.

From the standpoint of Kanthal curve, the company should identify and segment their customers so they can distinguish profitable from unprofitable customers. It is easy to conclude that the group with most profitable customers makes the company the largest profits and enterprise should retain them because they make a profit that is greater than 100% amount of profit, but what about the other two groups?

Those customers who belong to the group which approximates to zero, the company should stimulate with various marketing activities to cross in the most profitable group and not allow them to pass into the group of unprofitable. Group of unprofitable customers the company should try to switch, for the start, in the group of those whose profits approximate zero.

However, if the action fails, the company should leave them if they want to save the losses that they bring to them.

During this research, models used are:
- Linear Regression (LR-0-B-L)
- Multilayer Perceptron (MLP-1-B-L)
- Probabilistic Neural Network (PNN-0-N-N)
- Radial Basis Function (RBF-1-B-L)
- Generalized Feedforward (GFF-1-B-L)
- MPL with PCA (MLPPCA-1-B-L)
- Classification SVM (SVM-0-N-N)
- Multilayer Perceptron (MLP-2-B-L)

NeuroSolution training, selection and testing data presented in the following way shown in Tables 1.-3.:
Table 3 Testing – Performance Metrics

Multiple Layer Feedforward had the best performance with 77.60% accurate data in the training phase. Although in the training phase Classification SVM model captures accuracy of 93.59%, in the relevant test dataset Multilayer Perceptron has the highest prediction accuracy. The following is the Cross Validation process on the selection data set. In this part the network recorded the best performance of the Classification SVM model with 73.39% accuracy predicting observed customer segment. In the testing process, where the network should have learned to give the output variable, the Multilayer Perceptron model showed highest accuracy of 69.48%.

Summary of performance of the network with the best performance is presented in Table 4:

Table 4: Summary of the Performance Metrics of the best performing network

Table 4. shows the following. In the process of training data number was 951, of which the exact calculation obtained for 738, i.e. 77.60%, and incorrect calculation for them 213 or 22.40%. In the process of training data, data was taken into the sample data from 2009. and 2010. In the Cross Validation process, the number of data taken in the sample amounted to 511, of which exact calculation was made for 371 or 72.60%, and incorrect calculation for 140 or 27.40%. Data for the Cross Validation process was the data from 2011. In the testing process, in the sample was taken data from 2012., 829 of them. The exact number of test data was 576, i.e. 69.48%, while the number of incorrect test data was 253, i.e. 30.52%.

6 Conclusion

Classical methods of competition and gaining advantages in the global market is slowly becoming a thing of the past. The modern company has huge amount of information about the business environment and their own business. Companies must use mentioned data in a way to achieve competitive advantages, establish efficient processes and achieve optimal business results. The relationship with the customer makes only one, but a very important cog in the mentioned mosaic.

Enterprises of all sizes must strive the ideal of establishing close relations (1: 1) with its customers, understand each customer in particular and to take advantage of understanding them in the way to convince customers that it is better to do business with them than with the competition.

The calculation of the current profitability is a complex process. For company needs it is necessary to develop a general framework that will include all the necessary input variables and measures which they will show their profitability.

Basis is formed by linear measure costs and prices, and next to them there is a series of linear and non-linear variables that describe the features of customers of the company whose business with customer is analyzed.

Neural networks through various algorithms can extract and absorb the hidden knowledge in the existing data. This self-study based on data from the past, can generate indicators of different aspects of customers profitability through the application in problem domains such as clustering, association or classification data. Therefore, neural networks, combined with other methods of machine learning, make a good choice when choosing tools to generate a holistic view of the issue of the current customer profitability and to predict future of customer’s profitability. Due to the above authors choose neural networks to be the technique by which customer segmentation will be made.

The aim of the empirical research was to detect ability (or inability) of neural networks, based on the input data with the exception of the size of profit, and to classify them in the appropriate customer segment.

By calculating network obtained data which indicate that the Multilayer network was the most successful.
in the testing process, by providing accurate information in the 69.47% of cases.
What can be concluded from the obtained accuracy percentage of the network is that need further research, network training on a larger data set in order to come to a model that will provide a higher percentage of accuracy of the output.

There are many different methods of data mining like Cluster Analysis, Factor Analysis, Discriminant Function Analysis, Multidimensional Scaling, Log-linear Analysis, Canonical Correlation, Stepwise Linear and Nonlinear (e.g., Logit) Regression, Correspondence Analysis, Time Series Analysis, and Classification Trees, that can be used on the same dataset to compare results and to come to similar results.

In this paper is defined a small set of variables and indicators of model for prediction of customer profitability. Future research and work on a model should result in expanded set of indicators, and new methods of analysis that contribute most to achieving the ultimate goal of research.

Contribution of this work can be reflected in the detecting customers which are important for the company, so that company does not have to make losses and lose time with those who are bringing low profit or even losses. According to the above company can pay attention to those customers who are truly profitable for them and are bringing them profit and highly positive results. By making business on this way companies will have advantage over other companies that are not using customer segmentation but are dealing on the same way with all customers.

References: