Customer Segmentation Using Data Warehouse and Neural Networks

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

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

ZVONKO MERKAŠ, Ph.D.
Libertas Business School
CROATIA
Zvonko.Merkas@gmail.com

Abstract: The success of a company depends on its effective customer management. An intelligent company can achieve this goal by using modern information technology. This paper shows how data warehouse and neural networks can be useful in the process of predicting customer segments. The aim of this paper is to determine how neural networks, as a data mining method, are able to predict the belonging of each customer to a specific segment based on data from data warehouse. This paper presents a research on analysing data set of customers of the company which take care of production and distribution of nuts, grains and dried fruit. As a result there will be three groups of customers: a group containing the most profitable customers, a group of customers with a profit close to zero, and a group of the least profitable customers. Depending on where a specific customer belongs, a company will use different marketing activities in order to retain them, move them to a more profitable segment or ultimately reject them. The customer activity in the first quarter will be used to predict a segment in which a customer belongs at the end of the year. This will show the predictive ability of neural networks and how reliable they are in the process of customer segmentation using data from data warehouse.

Key-Words: customer, profitability, data warehouse, data mining, neural networks

1 Introduction

The aim of this paper is the use of data warehouse and neural networks in the process of customer segmentation.

The worldwide emergence of information revolution affects every type of business and industry. Despite a huge amount of information stored inside company information systems about customers and their transactions, the firm can rarely exploit its full potential in decision making and customer management. Striving for a success, firms are trying to find means for efficient analysis of these data. Customer management is one of the primary

concerns of a company. Today, as ever, companies must keep their focus on customer satisfaction and loyalty. Data warehouse can collect and transform millions of records for comprehensive analysis and provide tools that firms can and should use to understand the customer behaviour, to efficiently satisfy the customer needs and fulfil customer expectations, and, finally, to gain a competitive advantage.

Patterns in customer needs and habits are analyzed on sufficiently large sets of data. However, more the data, the more difficult becomes their collection and processing into timely and relevant information and knowledge. If performed without adequate IT tools, this process often requires a lot of competence, and, ultimately, generates high costs and hardly measurable results.

Implementation of data warehouse begins with collection, enhancement, and purification of the daily operations' data from different sources. Availability of "enhanced" data by using neural networks helps firms recognize and profit from new possibilities to strengthen customer relations, attract new prospects, and adapt to growth.

The interest in neural networks is 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 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. 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 have found an area of their application exactly in this segment of real world.

Neural networks are relatively easy to use. The 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 such a powerful mechanism is very wide. They are used in almost all 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 customers 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.

Finally, returns on investment in IT systems, such as data warehouse and neural networks, are increased customer loyalty, segmentation of profitable customers, attraction of new customers, and increased firm's market value.

This paper is structured as follows. First, the study objectives are introduced. Next, the short overview of Data Warehouse (Section 2) and OLAP and Data Mining is given (Section 3). Methodology of neural networks is presented in Section 4. The section 5 shows the general framework for customer segmentation, which includes data warehouse model and description of model itself, framework for concrete model presented in the paper, including input and output variables. The results are exposed in section 6. Finally, the last section outlines the conclusions and discusses some directions for further research.

2 Data Warehouse

Firms desperately need information and knowledge. For this purpose, firms are accumulating vast amount of data from disparate sources such as transaction systems, management support system, communication system, third-party agencies, Web, publications, research results, etc. Problems of capturing different types of structured and unstructured data relate to normalization, filtering, grouping, cleansing, and data enhancement. The solution is a concept known as a data warehouse.

The data warehouse is a manner for storing and presenting data in databases and supporting decision-making. Put more simply, the idea of data

warehousing is as follows: it is necessary to take the data out of the operational base and put them in analytical base (data warehouses) preparing them for demanding analyses, for data mining and the search of information and knowledge for effective decision-making.

A data warehouse includes data from the subject area that are necessary for decision-making in this area. [11] A data warehouse contains logically integrated data from various applications. Data in a warehouse are unchangeable from the point of view of processing. A warehouse is time oriented because it contains data that describe the phenomenon in a longer time period, which makes possible comparisons and forecasting. A warehouse is a set of data that ensures that questions made and analytic processing will lead to the acquisition of information and knowledge.

The original label that pre-dates the data warehouse is still the best description of what we are designing: a decision support system [15].

A large transactional system processes thousands or even millions of transactions per day. Most of the data are used only once or rarely. Extracted and/or aggregated data would be used more often. A specific usage of operational data can be increased by filtering or aggregating them into data warehouse. In other words, data should be selected, separated from data that is not useful and transformed according to the requirements of the data warehouse model. This is the way to make data in the data warehouse consistent (i.e. "one version of the truth"), which enables easier and more efficient access and use of corporate data.

A dimensional data model is most often used in data warehouse. The terms commonly used in this type of model are: dimensions (context) and facts (measures).

Dimensions are a category of information in data warehouse and through them business operations can be analyzed. For example, the dimension can be the time. Each data warehouse has the time dimension. The dimensions (often) have hierarchical structure. Hierarchy is the specification of levels that represents a relationship between different attributes within a dimension. For example, one possible hierarchy in the Time dimension is Year → Month → Day.

The facts are typically numeric values that can be aggregated, and dimensions are groups of hierarchies and descriptors that define the facts. For example, sales amount is a fact, cost is a fact, etc.; time, product, store, etc. are the dimensions.

The design of a dimensional database consists of four steps in a particular order:

1. Choose the business process to model
2. Declare the grain
3. Identify the dimensions
4. Identify the fact. [5]

In designing dimensional data models the most commonly used schema types are star schema or snowflake schema which depends on personal preference and business needs. The both schemas include fact tables and lookup (dimensional) tables.

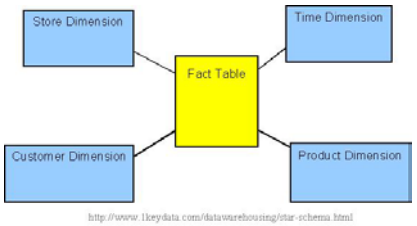

Fig. 1. Sample of star and schema [4]

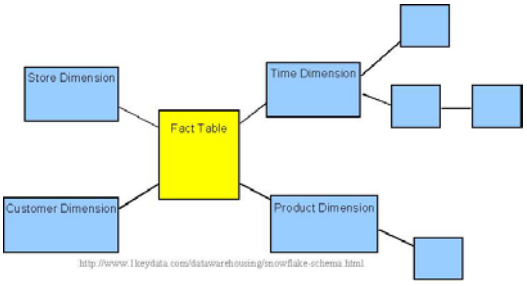

Fig. 2. Snowflake schema [4]

A fact table is a table that contains the measures of the interest. The lookup (dimensional) table provides the detailed information about the attributes of dimension. Fact tables connect to one or more lookup tables, but fact tables do not have direct relationships to one another. Dimensions and hierarchies are represented by lookup tables. Attributes are the non-key columns in the lookup tables.

The major characteristics of data warehousing are [25]:

- data are organized by a detailed subject, containing only information relevant for decision support,
- data in the warehouse are coded in a consistent manner,
- data are kept for several years so they can be used for trends forecasting, and comparisons over time, and
- once entered, data are not updated.

Benefits of data warehouses are most obvious in the companies with several computer platforms and versions and with many different data sources. The most important benefits companies are seeking from their DW efforts are: better business intelligence; reduced time to locate, access, and analyze information; consolidation of disparate information sources and strategic advantage over competitors.

3 OLAP and Data Mining

The next important question is how to extract information and knowledge from the data with issues in consistency, exactness, timeliness, and data complexity [2]. Using OLAP and Data Mining for knowledge discovery is a common approach to the problem. On-Line Analytical Processing and Data Mining are common methods for retrieving hidden knowledge from the data stored in a Data Warehouse [17].

On-line Analytical Processing enables manipulation and analysis of large amount of data, comparison of different types of data, complex computations and, most importantly, an intuitive graphical user interface (GUI) for presentation of results in various perspectives including drill-up and drill-down capabilities.

Data Mining is founded on algorithms for detection of unknown and unexpected patterns in large sets of data, clustering and segmenting of data and finding dependencies between multidimensional variables. The results of Data Mining analysis are presented graphically with the dominant and unexpected behavioural patterns enhanced.

In contrast to OLAP being retrospective in nature [25], data mining provides prospective knowledge discovery. Gartner Group defined a data mining as the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories, using recognition technologies as well as statistical and mathematical techniques.

Data mining technology discovers hidden trends and patterns in large volumes of data. A significant distinction between data mining and other analytical tools is in the approach they use in exploring the relationships among the data. The analytical tools usually support a verification approach, in which the user hypothesizes about data interrelationships are verified or refuted. This approach relies on the intuition of the analyst to pose the question and his or her ability to refine the analysis based on the results of potentially complex queries against a database.

Data mining uses discovery-based approaches in which pattern matching, clustering, and other algorithms are used to determine the significant relationships and correlation among data. Data mining algorithms can look at numerous multidimensional data relationships concurrently, highlighting those that are dominant or exceptional.

The data mining includes: association, classification, estimation, prediction, affinity grouping, clustering, description and knowledge discovery.

There are several techniques and algorithms used in data mining tools, from statistic analysis, through rule induction and nearest neighbor algorithms, to neural networks and genetic algorithms [14]. Data mining enables users to discover knowledge and provides them with greater depth and understanding of data than ad hoc querying and using of OLAP applications.

Applications for Data Warehousing, On-line Analytical Processing, and Data Mining are being widely adopted in modern companies to provide timely answers to many questions which previously required costly and lengthy programming and batch processing [3].

4 Neural networks

A neural network, as data mining techniques, 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. [8]
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 w_{K1}, w_{K2}, ..., w_{KN} 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.

A weighting function of network is its key elements and represents 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 pass on 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 used are the following four functions: Unit step (threshold), Sigmoid, Piecewise Linear and Gaussian. For the purpose of this study a non-linear sigmoid function is used whose formula is: [22]

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

Since the beginning of their presence in science, neural networks have been investigated with two different approaches. First, the biological approach, investigates the neural network as a simplified simulation of the human brain and uses them to test hypotheses about the functioning of the human brain. Another approach treats the neural network as technology systems for complex information processing. [16]

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 are used, so network learns on the basis of system behaviour 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. [24]

4.1 Backpropagation algorithm

The best known and simplest learning algorithm of neural networks is backpropagation algorithm. [20]
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, the 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. [19]

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. [19]

4.2. Previous researches

Successful application of neural networks is present in many areas, such as 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% [21]

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%. [1]

A 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. [9]

Young researcher from Google, Jim Gao, published his research in the paper called “Machine Learning Applications for Data Centre Optimization”. As a conclusion he implied that accelerating growth in DC (Data Centres) 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 centre) PUE (Power Usage Effectiveness) within 0.004 +/- 0.005, approximately 0.4% error for a PUE of 1.1. The 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. [18]

5 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.

Different customers have different characteristics such as preference, values, profit, etc.. Customer segmentation theory proposes that groups of customers with similar needs and purchasing behaviours are likely to demonstrate a more homogeneous response to marketing programs. With proper market segmentation, enterprises can arrange the right products, services and resources to each target customer cluster and build a close relationship with them. Market segmentation has consequently been regarded as one of the most critical elements in achieving successful modern marketing and customer relationship management. [26]

5.1. Data Warehouse modeling

In our model, company wants to have better understanding of customer purchases so they can segment them properly.

This data will allow us to analyze what, when and how often are customers purchasing and by that divide them into specific segments.

Granularity can reflect single transaction, daily snapshoot or monthly snapshoot. In this research, we can observe information on all three mentioned levels. [15]

Dimensions used in this model are following:

- Date dimension
- Store dimension
- Product dimension
- Customer dimension
- Promotion dimension

Fact table is Sales Transaction.

Dimensional model will be shown by star schema.
A star schema model can be depicted as a simple star: a central table contains fact data and multiple tables radiate out from it, connected by the primary and foreign keys of the database. In a star schema implementation, Warehouse Builder stores the dimension data in a single table or view for all the dimension levels.

For example, if you implement the Product dimension using a star schema, Warehouse Builder uses a single table to implement all the levels in the dimension, as shown in the screenshot. The attributes in all the levels are mapped to different columns in a single table called PRODUCT. [23]

Star Schema will be shown in Fig.3 Star Schema.

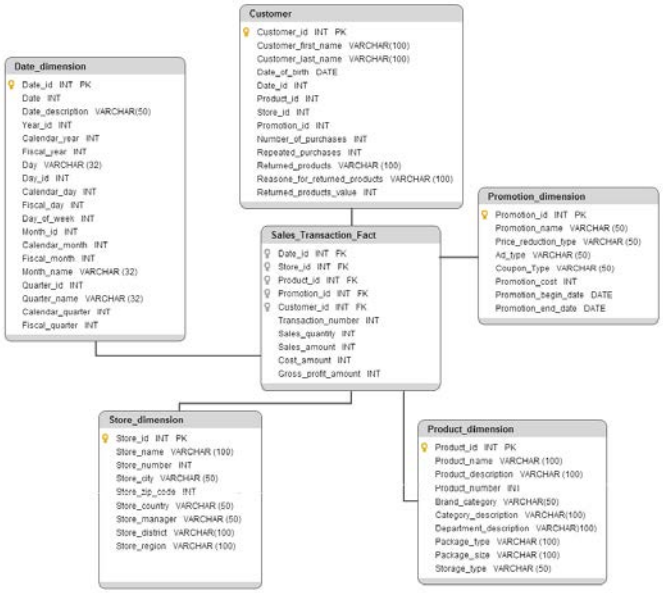

Fig.3. Star Schema

5.2 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. [10]

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.

5.3 Modeling a neural network for customer segmentation

The following are detailed descriptions of input and output variables.

5.3.1 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 it 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.

5.3.2 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 there 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. 4. [13]

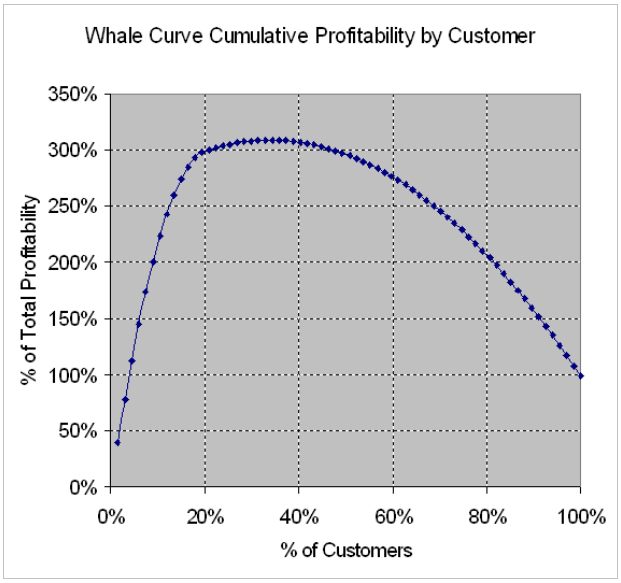

Fig. 4. Kanthal curve [13]

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. [1] 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 of 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. [12]

6 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 step into the most profitable group and not allow them to move to the group of the unprofitable. As to the group of unprofitable customers, the company should try to switch them, for a starter, to the group of those whose profits approximate zero.
However, if the action fails, the company should leave them if they want to avoid 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.:

Performance Metrics

	Training		
Model Name	**MSE**	**r**	**Correct**
LR-0-B-L (Linear Regression)	0,201644	0,414019	73,50%
MLP-1-B-L (Multilayer Perceptron)	0,161866	0,580015	77,60%
PNN-0-N-N (Probabilistic Neural Network)	0,183348	0,518364	74,97%
RBF-1-B-L (Radial Basis Function)	0,174029	0,533762	74,34%
GFF-1-B-L (Generalized Feedforward)	0,158809	0,589908	79,07%
MLPPCA-1-B-L (MLP with PCA)	0,172157	0,540923	75,60%
SVM-0-N-N (Classification SVM)	0,068887	0,873967	93,59%
MLP-2-B-L (Multilayer Perceptron)	0,16539	0,577695	77,71%

Table 1. Training – Performance Metrics

Performance Metrics

	Cross Validation		
Model Name	**MSE**	**r**	**Correct**
LR-0-B-L (Linear Regression)	0,212277	0,36779	66,14%
MLP-1-B-L (Multilayer Perceptron)	0,175136	0,509377	72,60%
PNN-0-N-N (Probabilistic Neural Network)	0,187541	0,453195	68,10%
RBF-1-B-L (Radial Basis Function)	0,192361	0,447191	70,06%
GFF-1-B-L (Generalized Feedforward)	0,176504	0,50781	71,62%
MLPPCA-1-B-L (MLP with PCA)	0,186629	0,467738	71,23%
SVM-0-N-N (Classification SVM)	0,186204	0,487235	73,39%
MLP-2-B-L (Multilayer Perceptron)	0,17344	0,511663	72,80%

Table 2. Cross Validation – Performance Metrics

Performance Metrics

	Testing		
Model Name	**MSE**	**r**	**Correct**
LR-0-B-L (Linear Regression)	0,429708	0,216106	67,91%
MLP-1-B-L (Multilayer Perceptron)	0,208346	0,415601	69,48%
PNN-0-N-N (Probabilistic Neural Network)	0,203883	0,428764	67,91%
RBF-1-B-L (Radial Basis Function)	0,207299	0,425618	67,91%
GFF-1-B-L (Generalized Feedforward)	0,20983	0,422723	67,31%
MLPPCA-1-B-L (MLP with PCA)	0,207594	0,430789	66,59%
SVM-0-N-N (Classification SVM)	0,215286	0,377953	64,54%
MLP-2-B-L (Multilayer Perceptron)	0,220765	0,404749	68,76%

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 re-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%.

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

Performance Metrics

	Training	Cross Val.	Testing
# of Rows	951	511	829
MSE	0,161866	0,175136	0,208346
Correlation (r)	0,580015	0,509377	0,415601
# Correct	738	371	576
# Incorrect	213	140	253
% Correct	77,60%	72,60%	69,48%

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, the sample contains 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%.

7 Conclusion

Conditions of business operations are marked by growing complexity and dynamics. As a result, an increasingly large number of business events in the firm and in its environment are generated, data about which have to be encompassed and stored in the system. Effective customer management required specific information and knowledge about all operational details. The firm's business processes with customers have to be recorded and appropriately stored. This historical data have to be accessible for analysis and knowledge extraction. In addition to this is the requirement for a more flexible approach. The solution is to create data warehouse.

In this paper a data warehouse contains selected data and information that have to be analysed and given

shape for the purposes of customer segmentation. Sales data are considered through five dimensions: time (date), store, product, customer and promotion. Such multidimensional structure of data, comprise a dimension data structure, and through the unification of them a model of a data warehouse for managing customer is created.

When data warehouse is created we could extract knowledge from the data using information technology such as neural network.

Classical methods of competition and gaining advantages in the global market are 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 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. The base 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. The aim of the empirical research was to detect ability (or inability) of neural networks, based on the input data in data warehouse with the exception of the size of profit, and to classify them in the appropriate customer segment.

By calculating network data are obtained 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 further research is needed, a 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 that can be used on the same dataset to compare results and to come to similar results. In this paper a small set of variables and indicators of model for prediction of customer profitability is defined. A 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 detecting customers who 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, a 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 in this way companies will have advantage over other companies that are not using customer segmentation but are dealing in the same way with all customers.

References:

[1] Coats, P., Fant, L., *Recognizing financial distress patterns using a neural network tool*, Financial Management, 1993., Vol. 22, pp. 142–155

[2] Ćurko K., Bosilj Vukšić, V., Lovrić, A., *Business Process Management Systems and Business Intelligence Systems as support of Knowledge Management*, Proceedings of the 8th WSEAS International Conference on Management, Marketing and Finances (MMF'10), Penang, Malaysia, March 23-25, 2010, pp 53-58

[3] Ćurko K., Varga M., Lončar A., *The support of Business Intelligence Technology in Process and Business Engineering*, International Journal of Computers, Issue 1, Volume 2, 2008, pp 8-14

[4] Data Warehousing Concepts [online] Available at: http://www.1keydata.com/datawarehousing/dimensional.html [25.03.2015]

[5] Dimensional Modeling – Wikipedia, free encyclopedia [online] Available at:

http://en.wikipedia.org/wiki/Dimensional_modeling [20.03.2015]

[6] Educational Repository for Intelligent Systems. [online] Available at: http://eris.foi.hr/11neuronske/nn-primjena.html. [15.01.2015.]

[7] Gheorghe, R., *Learning perceptron neural network with backpropagation algorithm*, Economic Computation & Economic Cybernetics Studies & Research, 2010., Vol.4, Issue 4, pp. 10

[8] Haykin, S., *Neural networks: A Comprehensive Foundation*, Prentice Hall, New Jersey, 1994. pp.2

[9] *How Google Uses Machine Learning and Neural Networks to Optimize Data Centers*. Available at: http://techcrunch.com/2014/05/28/how-google-uses-machine-learning-and-neural-networks-to-optimize-data-centers/ [18.01.2015.]

[10] Implementing ABC Accounting, Customer Profitability, and Product-Line Analysis in a Distribution Business [online] Available at: http://kczx.gzhu.edu.cn/course_center/files_upload/template/6C473DC0-9248-4A50-A2B2-04741AAD2F02/COLUMN_16/file3/IMPLEMENTING%20ACTIVITY-BASED%20COST%20ACCOUNTING.pdf [18.01.2015.]

[11] Inmon, W. H., *Building the Data Warehouse*, John Wiley & Sons, New York, 1996.

[12] Kaplan, R.S., Anderson, S.R., *Time-driven Activity-Based Costing,* Harvard Business School Press, 2007., pp.246

[13] Kaplan, R.S., Narayanan, V.G., *Customer Profitability Measurement and Management,* Harvard Business School, Boston, Massachusetts, 2001., pp. 4

[14] Kennedy R., Lee Y., Van Roy B., Reed C. D., Lippmann R. P., *Solving Data Mining Problems through Pattern Recognition*, Prentice Hall, 1998.

[15] Kimball, R., Ross M., *The Data Warehouse Toolkit – The Complete Guide to Dimensional Modeling*, Wiley. 2002.

[16] Klicek, B., Oreski, D., Oreski, S., *Bayesian and Neural Networks for Customer Satisfaction Optimization*, Proceedings of the 5th International Conference on Theoretical and Applied Mechanics (TAM'14), Lisabon, Portugal, October 30 – November 1, 2014, pp.119-129

[17] Laudon C. K, Laudon, P. J., *Management Information Systems: Managing the Digital Firm* 9-th ed., Pearson Prentice Hall. 2006.

[18] Machine Learning Applications for Data Center Optimization [online] Available at: http://static.googleusercontent.com/media/research.google.com/hr//pubs/archive/42542.pdf [18.01.2015.]

[19] Principe, J., Euliano, N., Lefebvre, W., *Neural and Adaptive Systems: Fundamentals through Simulations.* John Wiley & Sons, New York, 2000., pp. 150

[20] Ruxanda, G., *Learning Perceptron Neural Network with Backpropagation Algorithm,* Economic Computation & Economic Cybernetics Studies & Research, 2010., pp. 4

[21] Salchenberger, L., Cinar, E., Lash, N., *Neural networks: A new tool for predicting thrift failures,* Decision Sciences, 1992., vol. 23, pp. 899–916

[22] Song, Z., Xu, J., *Bifurcation and chaos analysis for a delayed two-neural network with a variation slope ratio in the activation function,* World Scientific Publishing Company. 2010., pp. 2

[23] Star and Snowflake Schemas [online] Available at: http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/10g/r2/owb/owb10gr2_gs/owb/lesson3/starandsnowflake.htm [22.03.2015]

[24] Turban, E., Aronson, J., E., *Decision support systems and intelligent system*, Prentice Hall, New Jersey, 2005. pp. 547

[25] Turban, E., McLean, E., Wetherbe, J., *Information Technology for Management: Making Connections for Strategic Advantage*, 2nd Edition, John Wiley & Sons, 1999.

[26] Yan, X., Li, Y., *Customer Segmentation based on Neural Network with Clustering Technique*, Proceedings of the 5th Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006, pp.265-268