

# A DATA MINING BASED APPROACH TO A FIRM'S MARKETING CHANNEL

*ESRA Kahya Özyirmidokuz*

*assistant professor / Ph.D., Computer Technologies Department, Erciyes University, Kayseri, Turkey, esrakahya@erciyes.edu.tr*

*KUMRU Uyar*

*assistant professor / Ph.D., Marketing and Production Department, Nuh Naci Yazgan University, Kayseri, Turkey, kumruuyar@hotmail.com*

*MUSTAFA HAKAN Özyirmidokuz*

*Bosch Thermotechnic, Ankara, Turkey, hakan.ozyirmidokuz@tr.bosch.com*

**Abstract:** *Firms need to collect and analyze marketing data to have a competitive advantage in the sector. The aim of this research is to extract knowledge from an international firm's marketing channel to improve the efficiency of the marketing system. Cross Industry Standard Process for Data Mining (CRISP-DM) is used to analyze the survey data. Data are clustered by applying a Kohonen Self Organizing Map (SOM) to reduce the attributes. Anomaly detection analysis is applied. We generate a C5.0 Decision Tree (DT) model used for predicting the marketing channel firms' complaints with very high accuracy. Decision rules are also extracted.*

**Key words:** Knowledge discovery in databases, data mining, DT, rule induction, marketing channel complaints

**JEL classification:**

## 1. Introduction

A marketing channel which delivers a firm's products and services to consumers, is one of the critical success factors of a marketing system to achieve the marketing objectives. If a firm doesn't know its marketing channel well, it can't manage it. Firstly, the firm must collect data from its marketing channel. The firm need to analyze data seriously to have a competitive advantage in the sector.

Data Mining (DM), which is the process of automatically searching large volumes of data to extract knowledge from data in a human-understandable structure, helps analysts recognize relationships within data.

Applying DM techniques to marketing data is extremely useful to find interesting, previously unknown, hidden patterns, which can then be better defined, in massive datasets. In this manner, DM helps to find important knowledge from marketing channel. The achieved knowledge has a strategic importance of competition and improvement of marketing and production for the firm. Because achieved knowledge can help to improve the communication between marketing channel and the firm by better controlling the processes, and by knowing the details about them.

One important type of knowledge that can be obtained from data mining is the decision tree (DT), which is constructed from existing data to classify future data. DTs are an effective method of classifying data set entries and can provide good decision support capabilities. DTs have several advantages over other data mining methods, including being human- interpretable, well-organized, computationally inexpensive, and capable of dealing with noisy data. Due to these merits, DTs are probably the most popular mining method (Yen-Liang Chen, Hsiao-Wei Hu, Kwei Tang, 2009). There have been numerous studies in marketing which uses decision trees (DTs) (Crone, Lessmann, Stahlbock, 2006; Lee, Lee, Park; 2007; Yada, Ip, Katoh, 2007; Abrahams et al., 2009; Kim, Timothy, Hwang, 2011; Olson, Chae, 2012; Duchessi, Lauría, 2013; Lei, Moon, 2015; You et al., 2015).

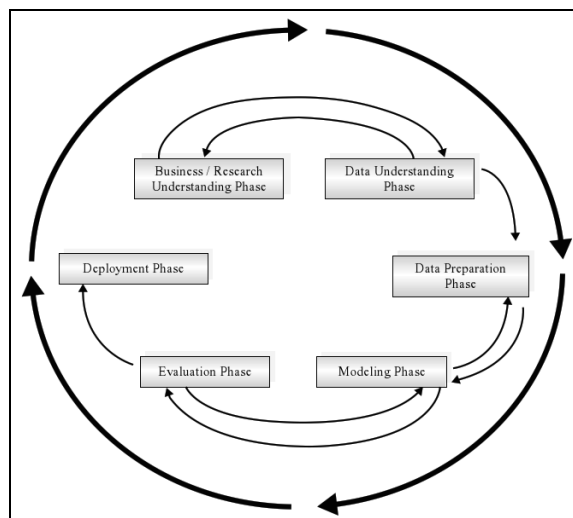
Among the data mining techniques, cluster analysis helps in classification of data. Cluster analysis seeks to maximize between-group variances and minimize within-group variances, including both hierarchical and non-hierarchical methods (Berson, Smith, Thearling, 2001).

In the literature, Kohonen's SOM is one of the techniques used for dimension reduction. Malone et al. (2005) demonstrated a trained SOM (Self-Organizing Map) which could provide initial information for extracting rules that describe cluster boundaries. Wang, Abrahamb and Smith (2005) used a SOM for pattern analysis and a fuzzy inference system to capture the chaotic trend to provide short-term (hourly) and long-term (daily) Web traffic trend predictions. Fessant, François and Clerot (2007) used Kohonen

SOMs and they showed how the mining of network measurement data can reveal the usage patterns of ADSL customers. Maiorana (2009) proposed a feature selection method based on a clustering algorithm belonging to the Kohonen SOM family. Gomez-Carracedo et al. (2010) applied Kohonen SOMs to perform pattern recognition in four datasets of roadside soil samples obtained in four sampling seasons over a one year period. They used CART as an objective variable selection step before the SOM grouping. Eshghi, Haughton, Regrand, Skaletsky and Woolford (2011), compared three clustering techniques: traditional clustering methods, Kohonen maps and latent class models. Nohuddin et.al. (2012), introduced a technique that uses frequent pattern mining and SOM techniques to identify, group and analyzed trends in sequences of time stamped social networks so as to identify interesting trends. In recent years, Kohonen SOM method is used in marketing (Gómez-Pérez et al., 2009; Lee, Park, 2005; Mostafa, El-Masry et al., 2012; Dragomir, Dragomir, Radulescu, 2014, Cuadros, Domínguez, 2014).

In this research, we use CRISP-DM which was developed in 1996 by analysts representing DaimlerChrysler, SPSS, and NCR. CRISP provides a nonproprietary and freely available standard process for fitting DM into the general problem-solving strategy of a business or research unit. According to CRISP-DM which is shown in Figure, a given DM project has a life cycle consisting of six phases. The phase sequence is adaptive. That is, the next phase in the sequence often depends on the outcomes associated with the preceding phase (Larose, 2005).

**Figure 1: CRISP-DM process**



Source: (Larose, 2005)

In the data preparation phase, firstly, we applied anomaly detection analysis which is used to reduce the number of records in a database. In addition, we determine the important features with Kohonen SOMs as a dimension-reduction technique to reduce the features. Although many different techniques could have been used in this study, e.g., PCA, factor analysis, and attribute relevance analysis, we used KNs in clustering due to the strength of Kohonen maps that lies in their ability to model non-linear relationships between data. The Kohonen map is one of the major unsupervised artificial neural network models. Kohonen Maps are useful tools for DM models with large data sets. High-dimensional data are projected to a lower dimension representation scheme that can be easily understood. In addition, Kohonen Maps can be used to process qualitative variables as well as quantitative ones (Tuffer, 2011).

DT induction is the modeling step in the forecasting process and consists of the determination of the DTs to the generalization of previously defined network behavior classes. DTs are easier to understand and they offer an acceptable level of accuracy. Several advantages of the DTs as a classification tool appear in the literature (Rokach and Maimon, 2007). In this research, a C5.0 DT is applied to the preprocessed data in order to predict the marketing channel firms' complaints about their international firms. Decision rules are extracted. The managers of the firm can easily understand these rules and predict future marketing channel firm's behaviour.

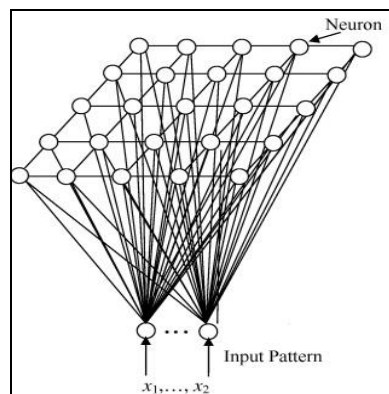
The paper is organized as follows. The Kohonen Network (KN) is introduced in Section 2. Section 3 gives the details of the application. The conclusions are drawn in Section 4.

## 2. Kohonen Networks

The identification of information, or patterns, in large subsets of data is a property of the fields of data-mining and feature extraction. Unsupervised learning techniques are a subset of these fields which enable the identification and grouping of patterns without having seen that pattern before or having its key characteristics described; to do this a similarity measure is defined and the groups are clustered together into a lower dimensional space. Self- Organizing Maps (SOM) are one such technique which enable the mapping of data with a large feature set into two-dimensional space. Moreover, SOMs allow the visual understanding of data structures. These can then be used as an aid in identifying and classifying anomalies in the datasets (Palomoa et al., 2012).

The goal of SOMs is to convert a complex high-dimensional input signal into a simpler low-dimensional discrete map. Thus, SOMs are very suitable for cluster analysis, when the underlying hidden patterns among records and fields are sought. SOMs structure the output nodes into clusters of nodes, where nodes in closer proximity are more similar to each other than other nodes that are farther apart. A typical SOM architecture is shown in Figure 1 (Larose, 2005).

**Figure 2: Topology of a simple SOM**



Source: (Larose, 2005)

KNs are a type of SOM, which itself represents a special class of neural networks. KNs can be considered a non-hierarchical method of cluster analysis. As non-hierarchical methods of clustering, they assign an input vector to the nearest cluster, on the basis of a predetermined distance function but they try to preserve a degree of dependence among the clusters by introducing a distance between them. Consequently, each output neuron has its own neighborhood, expressed in terms of a distance matrix. The output neurons are characterized by a distance function between them, described using the configuration of the nodes in a uni-dimensional or bidimensional space (Giudici, 2003).

## 3. Application

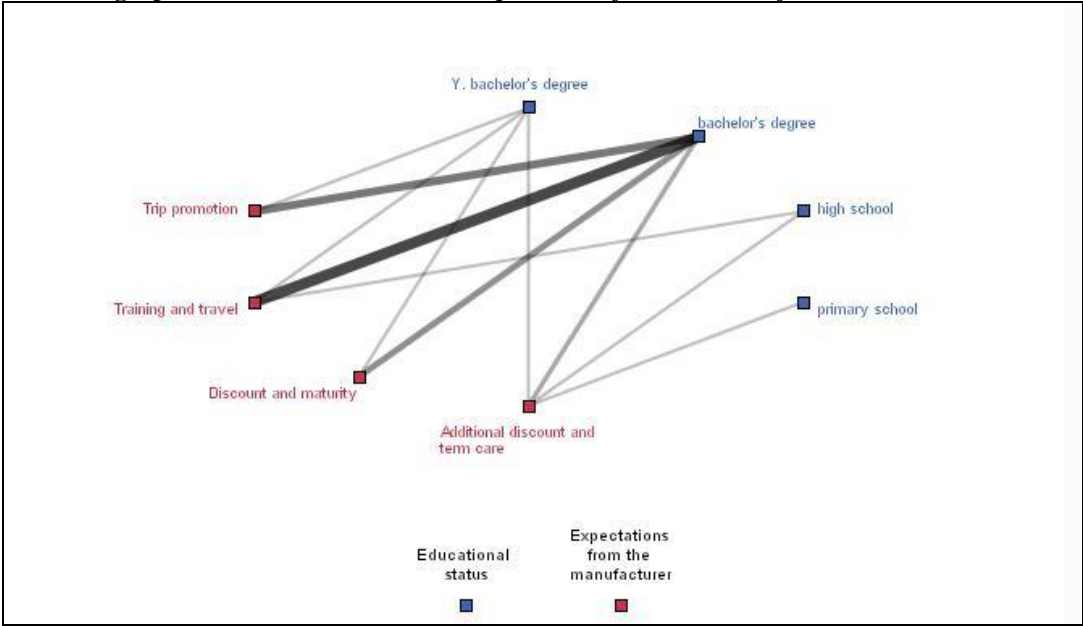
### 3.1. Understanding data

Data is collected manually from 300 firms of an international firm's marketing channel by one-to-one surveys. The collected surveys are converted to a data matrix which includes 20 features for each firm including education, sex, age, marital status, giro, the facility preferences provided by the firm, profession, number of trade marks, etc. The sex property is eliminated manually because all the marketing channel firms' owners are all male except one of them. In this phase, the basic statistical analysis and some exploratory graphical analysis techniques are used to understand the data matrix.

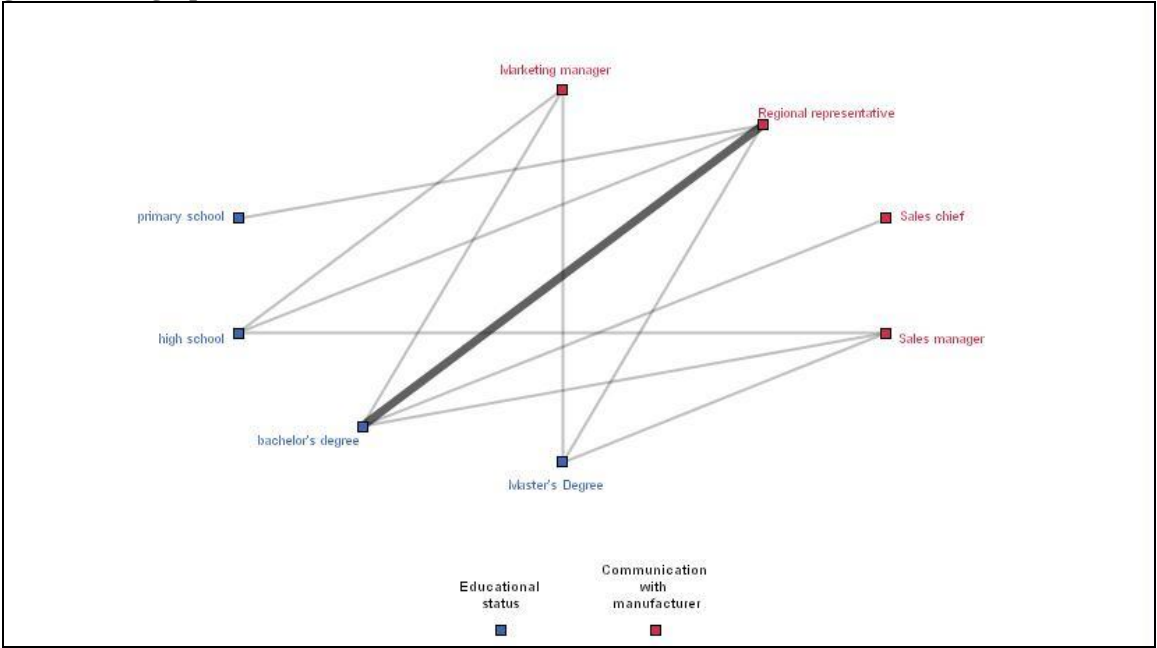
Web graphs are performed to understand the relationships between features. We will give two examples. Figure 3 illustrates the relation between educational status and the expectations of the marketing channel firms' owners from the manufacturer firm. We can easily see that the marketing channel firms' owners with *bachelor's degree* prefers *traveling and training* from the manufacturing

firm. But the owner with primary school degree doesn't prefer *traveling and training*. We can also see from the Figure 4 that marketing channel firms' owners with lower educational status prefers to communicate with regional representatives.

**Figure 3: Web graph of educational status and expectations from the manufacturer**



**Figure 4: Web graph of educational status and communication with manufacturer**



Although classical statistics work well in giving a picture of the current overall data matrix, they are unable to find hidden knowledge waiting to be discovered. In this manner, DM is used to find previously unknown and useful knowledge.

3.1. Data preparation phase

Anomalous records which show different behavior from the previously measured values in the data matrix must be detected. After we applied anomaly detection analysis, we detected two groups of records with 236 and 64. After anomaly detection analysis, only three records which are greater than anomaly index level (1.456) are eliminated from the data matrix.

It is known that feature selection improves model accuracy in knowledge discovery process. In this research, Kohonen SOM is used to choose a subset of input variables by eliminating features with little predictive information. Records are grouped by KN (Kohonen Network) so that records within a group or cluster tend to be similar to each other, and records in different groups are dissimilar. For the first 20 cycles, the neighborhood size was set at  $R=2$ , and the learning rate was set to decay linearly starting at  $\eta = 0.3$ . Then, for the next 150 cycles, the neighborhood size was reset to  $R=1$  while the learning rate was allowed to decay linearly from  $\eta = 0.3$  to  $\eta = 0.1$ . The neurons are organized into two layers, input layer and output layer. In the study, the input layer has 45 neurons and the output layer has 12 neurons. The self-organizing map identified 12 different clusters. The clusters and number of the records are illustrated in Table1. After Kohonen SOM model, we eliminate 11 features from the data matrix including sex and communication with manufacturer.

**Table 1: Table title (Times New Roman, 10 pt., bold, center)**

<b>x</b>	<b>y</b>	<b>Number of records</b>
0	0	42
0	1	22
0	2	48
1	0	22
1	1	4
1	2	14
2	0	15
2	1	6
2	2	26
3	0	47
3	1	2
3	2	49
<b>Total</b>		297

After preprocessing, there are 8 features in the data matrix as follows; *Educational status, Preferences about quality or price, Number of trade marks, Material status, Profession, Expectations from the manufacturer, Different sectors, Giro*. These features are the inputs of the DT model. The output of the model is *Complaints about the manufacturer*.

### 3.1. Decision tree induction

DTs are one of the popular methods that have been used for knowledge discovery in databases. Tree models can be defined as a recursive procedure, through which a set on statistical units are progressively divided into groups, according to a division of an explanatory variable to split and the choice of a splitting rule for such variable, which establishes how to partition the observations. The main result of a tree model is a final partition of the observations. To achieve this it is necessary to specify stopping criteria for the division process (Giudici, 2003). DT provide an easy to understand overview for users without a DM background with high classification accuracy. They also provide a tree model of the problem and various alternatives in an understandable format without explanation. The acquired knowledge is usually quite understandable and can be easily used to obtain a better understanding of the problem. In addition, DTs assist in making decisions with existing information. They have satisfactory performance even when the training data are highly uncertain. (Kahya-Özyirmidokuz et al., 2012). In the study, a C5.0 DT model is run to build a DT for predicting marketing channel's firms' complaints. The aim of the classification is to find similar data items which belong to the same class.

Modelling is performed with C5.0 which is one of the popular DT modelling algorithms, an extension of the earlier well known algorithm ID3. A C5.0 DT model which is acceptable with 10 tree

depths. The mean is 91.3 and the standard error is 1.4 for the cross validation. The relations and knowledge are acquired.

191 if-then rules are generated to express the process in English. The following examples illustrate some of the rules:

- if *Number of trade marks* > [ 4 ] and *Profession* is [ "Engineering" ] and *Expectations from manufacturer* is ["Additional discount and term care"] then *firm's complaint* is ["communicating problems"],
- if *Educational status* > [ "lisans" ] and *Profession* is [ "toptanci" ] and *Expectations from manufacturer* is ["Additional discount and term care"] and *Giro* <= [15000000] then *firm's complaint* is ["money collecting difficulties"],

*Stratified* ten-fold cross-validation accuracy evaluation was used to train and test the data matrix. The accuracy ratio of the model is 92.67 % which is shown in Table 1.

**Table 1: Model Accuracy**

	Number of records	Precent of records
<b>Correct records</b>	278	92.67%
<b>Wrong records</b>	22	7.33%
<b>Total records</b>	300	100%

#### 4. Conclusions

Today, marketing channel decisions are as important as the decisions companies make about the features and prices of products (Hitt, Black, Porter, 2009). In this research, we applied a DM framework and we presented a decision tree induction from marketing channel data to improve the efficiency of the marketing system. DM techniques were implemented to marketing servey data. We explored the use of different pre-processing and DM techniques including anomaly detection analysis, Kohonen SOM, and C5.0 DTs. This research included attribute reduction using KNs. A C5.0 DT which was used for the classification of the data set with 10 tree depths was generated. The accuracy ratio of the model was 92.67%. The DT model lays out the data matrix clearly so that all options can be explored. This acquired knowledge may be used to predict the future behaviors of the marketing channel firms. Data is processed into a manageable format, and decision rules are also generated. The DT model helps managers to understand the marketing channel firms. This research is also important to assess future complaints of the firms and to plan future marketing developments. If we evaluate the current marketing channel and plan for capacity needs we will achieve a better marketing system performance. Consequently, the knowledge will improve the performance of the marketing system.

Alternative DM techniques using artificial intelligence methods can be studied in future research to compare various approaches and to implement this framework.

#### ACKNOWLEDGMENT

This work was supported by the strategic grant POSDRU/159/1.5/S/133255, Project ID 133255 (2014), co-financed by the European Social Fund within the Sectorial Operational Program Human Resources Development 2007 – 2013.

This work was also supported by Erciyes University Research Fund, Project Number FBA-2014-5364.

#### 5. References

- Berson, A., Smith, S., Thearling, K. (2001) *Building data mining applications for CRM*. New York: McGraw-Hill.

- Chen, Y-L., Hu, H-W., Tang, K. (2009) Constructing a decision tree from data with hierarchical class labels, *Expert Systems with Applications*, 36, p. 4838–4847.
- Çiflikli, C., Kahya-Özyirmidokuz, E. (2012) Enhancing product quality of a process, *Industrial Management and Data Systems*, 112(8), p. 1181-1200.
- Crone, S.F., Lessmann, S., Stahlbock, R. (2006) The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing, *European Journal of Operational Research*, 173, p. 781–800.
- Cuadros, J., Domínguez, V.E. (2014) Customer segmentation model based on value generation for marketing strategies formulation Alvaro, *Estudios Gerenciales*, 30, p. 25–30.
- Dragomir, O.E., Dragomir, F., Radulescu, M. (2014) Matlab Application of Kohonen Self-Organizing Map to Classify Consumers' Load Profiles, 2nd International Conference on Information Technology and Quantitative Management, ITQM, *Procedia Computer Science*, 31, p. 474 – 479.
- Duchessi, P., Lauría, E.J.M. (2013) Decision tree models for profiling ski resorts' promotional and advertising strategies and the impact on sales, *Expert Systems with Applications*, 40, p. 5822–5829.
- Eshghi, A., Haughton, D., Regrand, P., Skaletsky, M., Woolford, S. (2011) Identifying groups: A comparison of methodologies, *Journal of Data Science*, 9, p.271-291.
- Fessant, F., François, J., Clerot, F. (2007) Characterizing ADSL customer behaviours by network traffic data-mining, *Annales des Télécommunications*, 62 (3-4), p. 350-368.
- Giudici, P. (2003) *Applied DM*, England: Wiley.
- Gomez-Carracedo, M. P., Andrade, J.M., Carrera, G.V.S.M., Aires-de-Sousa, J., Carlosena, A., Prada, D. (2010) Combining Kohonen neural networks and variable selection by classification trees to cluster road soil samples, *Chemometrics and Intelligent Laboratory Systems*, 102, p.20-34.
- Gómez-Pérez, G., Martín-Guerrero, J.D., Soria-Olivas, E., Balaguer-Ballester, E., Palomares, A., Casariego, N. (2009) Assigning discounts in a marketing campaign by using reinforcement learning and neural networks, *Expert Systems with Applications*, 36, p. 8022–8031.
- Hitt, M., Black, S., Porter, L. (2009) *Management*, 2nd ed., Upper Saddle River NJ: Prentice Hall.
- Kim, S.S., D.J. Timothy, Hwang, J. (2011) Understanding Japanese tourists' shopping preferences using the DT analysis method, *Tourism Management*, 3, p.544-554.
- Larose, D.T. (2005) *Discovering Knowledge in Data: An Introduction to DM*, USA: Wiley.
- Lee, J.H., Park, S.C. (2005) Intelligent profitable customers segmentation system based on business intelligence tools, *Expert Systems with Applications*, 29, p. 145–152.
- Lee, S., Lee, S., Park, Y. (2007) A prediction model for success of services in e-commerce using decision tree: E-customer's attitude towards online service, *Expert Systems with Applications*, 33, p. 572–581.
- Lei, N., Moon, S.K. (2015) A Decision Support System for market-driven product positioning and design, *Decision Support Systems*, 69, p. 82–91.
- Maiorana, F. (2009) Feature Selection with Kohonen Self Organizing Classification Algorithm, *World Academy of Science, Engineering and Technology*, 27, p. 998-1003.
- Malone, J., McGarry, K., Wermter, S., Bowerman, C. (2005) DM using rule extraction from Kohonen self-Organizing maps, *Neural Comput and Applic*, 15, p. 9-17.
- Mostafa, M.M., El-Masry, A.A. (2013) Citizens as consumers: Profiling e-government services' users in Egypt via data mining techniques, *International Journal of Information Management*, 33, p. 627– 641.
- Nohuddin, P.N.E., Coenen, F., Christley, R., Setzkorn, C., Patel, Y., Williams, S. (2012) Finding “interesting” trends in social networks using frequent pattern mining and SOMs, *Knowledge-Based Systems*, 29, p. 104–113.
- Olson, D.L., Chae, B.(K.) (2012) Direct marketing decision support through predictive customer response modeling, *Decision Support Systems*, 54, p. 443–451.

- Palomoa, E. J., North, J., Elizondo, D. , Luquea, R.M., Watson, T. (2012) Application of growing hierarchical SOM for visualization of network forensics traffic data, *Neural Networks*, 32, p. 275-284.
- Seret, A., Verbraken, T., Versailles, S., Baesens, B. (2012) A new SOM-based method for profile generation: Theory and an application in direct marketing, *European Journal of Operational Research*, 220, p. 199–209.
- Tuffery, S. (2011) *DM and Statistics for Decision Making*. USA: Wiley.
- Wang, X., Abrahamb, A., Smith, K.A. (2005) Intelligent web traffic mining and analysis, *Journal of Network and Computer Applications*, 28, p. 147–165.
- Yada, K., Ip, E., Katoh, N. (2007) Is this brand ephemeral? A multivariate tree-based decision analysis of new product sustainability, *Decision Support Systems*, 44, p. 223–234.
- You, Z., Si, Y-W., Zhang, D., Zeng, XX., Leung, S.C.H., Li, T. (2015) A decision-making framework for precision marketing, *Expert Systems with Applications*, 42, p. 3357–3367.