DE GRUYTER
OPEN

Data Mining as Support to Knowledge Management in Marketing

Marijana Zekić-Sušac, Adela Has
Faculty of Economics in Osijek, Croatia

Abstract

Background: Previous research has shown success of data mining methods in marketing. However, their integration in a knowledge management system is still not investigated enough. **Objectives:** The purpose of this paper is to suggest an integration of two data mining techniques: neural networks and association rules in marketing modeling that could serve as an input to knowledge management and produce better marketing decisions. **Methods/Approach:** Association rules and artificial neural networks are combined in a data mining component to discover patterns and customers' profiles in frequent item purchases. The results of data mining are used in a web-based knowledge management component to trigger ideas for new marketing strategies. The model is tested by an experimental research. **Results:** The results show that the suggested model could be efficiently used to recognize patterns in shopping behaviour and generate new marketing strategies. **Conclusions:** The scientific contribution lies in proposing an integrative data mining approach that could present support to knowledge management. The research could be useful to marketing and retail managers in improving the process of their decision making, as well as to researchers in the area of marketing modelling. Future studies should include more samples and other data mining techniques in order to test the model generalization ability.

Introduction

In recent years, a high level of information availability enabled by Internet technologies such as cloud, social networks, web 2.0, web 3.0, on one side, as well as a rapid development of methodological aspects such as data mining (DM), data warehousing, and business analytics on the other side, open the possibility to develop and implement specific parts of knowledge management systems in a way which is more approachable to decision makers.

DM includes implementation of various advanced statistical and machine learning methods to reveal hidden relationships from large amounts of data. Results and models generated by DM methods could serve as an efficient support for knowledge management (KM). In this paper, we suggest a model based on DM techniques, specifically association rules (ARs) and neural networks (NNs). ARs and NNs are used in an integrative way to discover some interesting patterns of customer behavior, specifically focused on items that are frequently bought together, and on the profiles of customers who buy those items. However, in previous research DM methods were mostly used individually, while the user is left without an integrative framework on how to combine them in order to reveal useful knowledge that could be a trigger of innovative strategies. This paper aims to address that gap by focusing on the question of how to effectively integrate neural networks and association rules in order to produce useful knowledge for marketing decision making.

The purpose of the paper is to propose an approach of integrating different techniques of data mining and to suggest its usage in knowledge management to produce better marketing decisions. The purpose will be realized through objectives: (1) to propose a new experimental approach of integrating neural networks and association rules in marketing modeling, (2) to assess predictive power of the combined techniques, and (3) to discuss implication of this approach in knowledge management in marketing.

ARs are selected due to its ability to discover hidden relationships among large amount of data primarily from transactional databases (Liao et al., 2009). One of the problems in which ARs can be applied is the market basket problem which assumes that there are a large number of products that can be purchased by a customer, either in a single transaction or over time in a sequence of transactions. NNs are selected due to its suitability for both classification and prediction type of problems such as a customer behaviour which can be observed as a binary or a multiple response classification problem, as well as a regression problem of predicting a continuous output. ARs are used in the first stage of the model to discover the most frequent items, while NNs are used in the second stage to discover the profile of customers who are likely to buy the most interesting items or sets of items identified in the first stage. Such methodological approach is then incorporated into a knowledge management system of a company therefore enabling continuous usage of DM as a knowledge management support.

Literature review

The overview of previous literature in the area of DM and KM provide some clear indications of their strong interconnections.

Kotler and Armstrong (2010) emphasized that marketing information should be used to gain customer insights and make better marketing decisions. The paper of Shaw et al. (2001) suggests a framework for KM in the context of marketing and concludes that systematic application of DM techniques can enhance the KM process and enable marketers to know their customers better, and improve customer service. Rygielski et al. (2002) examined the integration of customer relationship management (CRM) and DM as an important tool for achieving business competitive advantages trough identifying valuable customers, predict future behaviors, and enable firms to make proactive, knowledge-driven decisions. They integrated two DM techniques: Chi-square Automatic Interaction Detection (CHAID) and NNs. Heinrichs and Lim (2003) demonstrated a positive interaction effect between the DM tools and models application on strategic performance and inferred that with a proper use of web-based knowledge generation tools, the

business can achieve a significant competitive advantage. Javaheri et al. (2013) also showed a DM-based approach for target selection in response marketing. A unified theoretical framework for DM is given by Khan et al. (2013) in which they suggest clustering, classification trees and visualization techniques to support decision makers in marketing.

Previous research shows that rapid information technology development greatly contributed the development of KM in today's businesses and that integration of KM and DM mining could be a way to overcome the obstacles in efficient implementation of KM (Tsai, 2013). The integrative models of KM and DM are still not investigated enough in marketing decision support, which is a gap addressed in this paper.

Methodology

This section provides a brief overview of data and DM methods used in the research, as well as the description of suggested model for integrating DM based on ARs and NNs as a support in KM.

Data

Data from a transactional database of a retail store were used by a DM software tool (Statistica Datamining) to produce initial association rules that revealed customers' shopping behaviour. The following input variables were used to describe the purchase: *purchase date, receipt number, item code*, and *item name*. The available customer characteristics were: *gender, education level, marital status*, and *income category* as categorical variables, in addition to *number of kids, number of cars, age* and *home value* as continuous variables.

The dataset contained 14012 transactions which occurred during a one month period. After the pre-processing phase of data cleansing and filtering, where transactions containing some very rare items were excluded, 7006 transactions remained with 278 different items which have appeared on 3158 different customers' receipts. In order to provide groups of items for the hierarchical ARs, we have grouped together items which represent the same type of product, but of a different manufacturer, packing, brand, net weight, or volume. The grouping procedure resulted with 38 large groups of items.

Association rules methodology

Association rules have shown their success in discovering unknown relationships in data, thus providing the basis for decisions in marketing, retail, education, and other areas. They were first introduced by Agrawal *et al.* (1993) who proposed this method for market-basket analysis. According to Liu et al. (2001) ARs can be described as follows. If $I = \{item_1, item_2, ..., item_m\}$ is a set of items, and D is a set of transactions (the dataset), where each transaction d is a set of items such that $d \subseteq I$, an association rule is an implication of the form $X \rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule has a support s in D if s percent of transactions in D contains $X \cup Y$. The rule $X \rightarrow Y$ holds in the transaction set D with confidence c if c percent of transactions in D that support X also support Y. Given a set of transactions D (the dataset), the problem of mining ARs is to discover all relevant rules. In ARs, any item can appear on the left-hand-side (called body or antecedent) or the right-handside of a rule (called head or consequent).

The standard algorithm used in ARs is the apriori algorithm introduced by Agrawal, Srikant (1994). ARs in our experiments are generated by an improved algorithm

called the tree-building technique, which compresses a large database into a compact, Frequent-Pattern tree (FP-tree) structure (Lin et al, 2011). The advantage of this algorithm is in its speed, since it scans the whole database only once. It uses a divide-and-conquer approach, such that if first computes the frequent items and characterizes them in a tree called frequent-pattern tree. The FP-tree serves as a compressed database on which the AR mining is performed. Besides that, this algorithm does not require the candidate itemset generation, and is therefore more efficient than the apriori algorithm (Lin et al, 2011). Disadvantage of FP-tree algorithm is in generating a large number of conditional FP trees recursively as a procedure of mining.

ARs can be evaluated by different measures, usually divided to objective and subjective measures (Silberschatz, Tuzhilin, 1995). The basic objective measures are support s and confidence c (Chen et al., 2013). According to (Liao et al., 2009) for an association rule $X \rightarrow Y$, $s(X)$ or $s(X \cup Y)$ is used to represent the generality of the rule, and $c(X \rightarrow Y)$ is used to represent the reliability of the rule. Although in general, a rule with high generality and reliability is considered interesting, many authors emphasize that rules with low generality can sometimes have very high reliability and therefore, could be very interesting. In this paper, we combine objective and subjective approach, such that we use unexpectedness and actionability as subjective measures suggested by Silberschatz & Tuzhilin (1995), and also the confidence c as an objective measure, where we consider a rule with a minimal confidence of 40% as potentially interesting one. Silberschatz & Tuzhilin (1995) suggest a computational method for measuring unexpectedness based on the frequency of items (i.e. support), while in this paper we use heuristic estimations, where an expert manager estimates unexpectedness and actionability as binary values such that:

$$heuristic\ unexpectedness = \begin{cases} 0, & if\ the\ rule\ was\ not\ expected\ based\ on\ heuristics \\ 1, & if\ the\ rule\ was\ expected\ based\ on\ heuristics \end{cases} \quad (1)$$

$$heuristic\ actionability = \begin{cases} 0, & if\ user\ cannot\ realize\ some\ action \\ 1, & if\ user\ can\ realize\ some\ action \end{cases} \quad (2)$$

where by the term „user" we consider a sales or marketing department of the observed company, while the term „action" denotes any marketing activity that could be made for increasing the sale of an itemset which construct a rule, such as discount, paired advertising, or other. In our experiments, the support s was selected in a cross-validation procedure such that different values of s were tested in order to find the value that is the most suitable for each of the tested models. Minimum s used to produce the rules ranged from 3% to 50%.

Neural network methodology

Artificial neural network (NN) as a DM method has been successfully used for both regression and classification type of problems in different areas (Paliwal and Kumar, 2009). It aims to approximate the function between the input vector and the output, whereby it tests various linear and nonlinear functions in more layers of computation to achieve the minimum error or cost. Besides in predictive purposes, the NNs can be used for exploratory analysis to reveal the importance of predictors in a model. Some of the advantages of NNs are in its robustness, the ability to work with missing data, the ability to approximate any nonlinear mathematical function (Masters, 1995), while the main limitations are in the possibility to achieve a local instead of the global minimum, and in their instability regarding the change of sample structure.

The most common type of NN was tested in this research - the multilayer perceptron (MLP), a feed forward network that is able to use various algorithms to minimize the objective function, such as backpropagation, conjugate gradient, and other algorithms.

The input layer of a NN consists of n input units $x_i \in R$, $i=1,2,...,n$, and randomly determined initial weights w_i usually from the interval [-1,1]. Each unit in the hidden (middle) layer receives the weighted sum of all x_i values as the input. The output of the hidden layer denoted as y_c is computed by:

$$y_c = f\left(\sum_{i=1}^{n} w_i x_i\right) \tag{3}$$

where f is the activation function selected by the user, which can be logistic, tangent hyperbolic, exponential, linear, step or other (Masters, 1995). The most common, logistic (i.e. sigmoid) activation function is computed according to:

$$f(x_i) = \frac{1}{1+e^{g \cdot x_i}} \tag{4}$$

where g is the parameter defining the gradient of the function. The output is in the interval [0,1]. The output of a NN is compared to the actual output y_a, and the local error ε is computed. The error is then used to adjust the weights of the input vector according to a learning rule, usually the Delta rule (Masters, 1995). The above process is repeated in a number of iterations (epochs), where the gradient descent or other algorithm is used to minimize the error. In order to produce probabilities in the output layer, a softmax activation function is added for classification purposes.
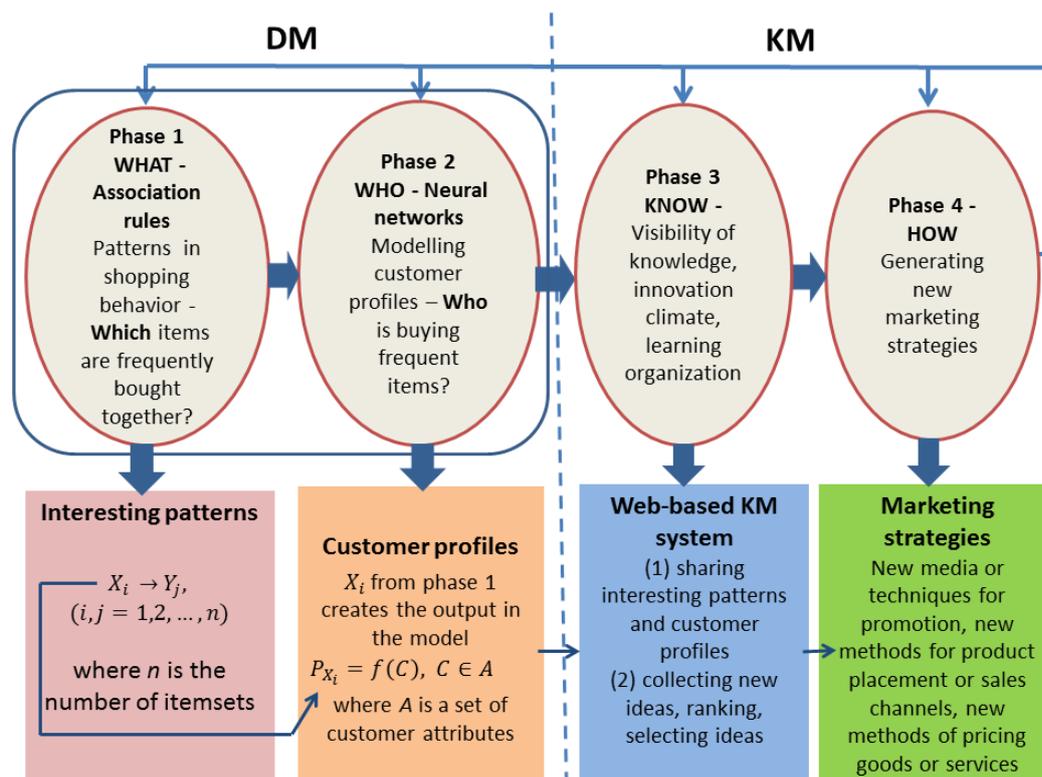
The output of the NN model created for the purpose of identifying customer profiles consists of a binary variable (valued as 1 for the existence of purchase, and 0 for the absence of purchase of a specific item or a set of items in a store identified as interesting by ARs). In order to find the most efficient NN model, extensive tests were performed by varying learning algorithms (backpropagation and conjugate gradient), activation functions (sigmoid, tangens hyperbolic, exponential, and linear), the number of units in the hidden layer (from 1 to 20). A crossvalidation procedure was used to optimize the training time, as well as to find the optimal number of hidden units. The maximum number of training epochs was set to 1000. After the learning and testing phases, the NN model can be used to reveal which of the input characteristics of customers is important to predict its purchase. This is enabled by a sensitivity analysis, which is performed on the out-of-sample test data, resulting in the set of important predictors. The sensitivity analysis is performed in a way that the program changes the value of an input variable by a randomly selected percentage value (in the range of +-5%), remaining all other input variables the same, and observes the change of the model error. The sensitivity coefficient of each input variable presents the ratio of the average model error with changes of an examined input variable in relation to model error without changes of an examined input variable. The variable whose sensitivity ratio is above or equal 1 is found as an important predictor that improves the model, while the variable whose sensitivity ratio is below 1 is found not to be important for the model. Upon the results of the sensitivity analysis, a user is able to extract the important predictors, and also to do a-posteriori analysis of values of the predictors that lead to the purchase.

Integration of DM and KM (the DataKnow model)

With the aim to provide more comprehensive description of integrating ARs and NNs into KM in marketing, we propose a model named the DataKnow, which is presented in Figure 1.

Figure 1
DataKnow Model of Integrating DM and KM in Marketing

Source: Authors

The suggested DataKnow model consists of two interconnected components: Data mining (DM), and Knowledge management (KM). The methodological activities in those components are represented in four phases. The activities are illustrated in ellipses in Figure 1, while their results, i.e. effects are represented in rectangles.

The **DM component** consists of two phases. The Phase 1 or "**WHAT**" phase, aims to discover sets of items that are frequently bought together. Methods of DM appropriate for such task are ARs and related methods such as link analysis and sequence analysis. They all provide patterns in the form $X \rightarrow Y$ (If X then Y), where $X \subset I$, $Y \subset I$ and $I = \{item_1, item_2, ..., item_m\}$ is a set of items in a store. This activity is to be conducted by a business analyst or an expert in the area of data mining. The source of data for this component can be a local transactional database or a web-based source that captures shopping behaviour in a web shop. The result of this first phase is a number of patterns (i.e. rules) that are generated by a DM tool. Due to a large size of transactional databases, the number of generated rules is usually also large, and requires additional filtering by using objective measures support s and confidence c, and subjective measures of unexpectedness and actionability estimated by heuristic expert knowledge to select the most interesting rules. The result of this component is a small set of interesting rules that are unexpected and confident, in the form:

$$X_i \rightarrow Y_j, \ (i, j = 1, 2, \dots, n) \tag{5}$$

where *n* is the number of the extracted itemsets. In order to enhance the sale of a consequent itemset Y_j it is crucial to see which customers buy their antecedents X_i. Therefore, we suggest that the antecedents X_i are forwarded to the next phase to model the purchase of X_i. In that way the ARs are used as a pre-modelling technique to determine the output form of the NN model.

The results of the first phase are forwarded to the second component: **Phase 2** or **"WHO"** component which is aimed to answer the question of "Who is buying frequent itemsets that are identified in Phase 1?" i.e. to identify the profiles of customers who are likely to buy important itemsets. In order to identify customers' profile, we suggest NN methodology. The NN model formulation adopted for identifying the profile of customers who bought frequent itemsets is:

$$P_{X_i} = f(C), \ C \in A \tag{6}$$

where P_{X_i} is the binary output denoting *Purchase of itemset* X_i, C is a set of customer characteristics used as the set of input variables, and A is the total set of customer attributes available in the dataset.

The procedure of NN modelling in this phase can be described by the following algorithm:

(1) Open a dataset including data on purchase transaction, customer characteristics, such as age, gender, education level, home ownership, number of cars, number of kids, and other available descriptive variables, as well as some behavior variables (such as number of previous purchases, time between purchases, etc.).
(2) Select customer descriptive and behavioral variables as input variables in the NN model.
(3) From each AR extracted in Phase 1, take the antecedent X_i.
(4) Use P_{X_i} as the output variable in the NN model, such that it has a binary response (0 if a customer *i* has bought the itemset X_i, and 0 otherwise).
(5) Design the NN architecture (number of hidden units, activation functions, number of epochs, subsampling, and the objective measure – classification rate or other) and run the NN model (training, testing and validation).
(6) Observe the NN accuracy on the validation subsample. If the accuracy is satisfactory (depending on the user needs), do the sensitivity analysis.
(7) Select important predictors of purchase upon sensitivity analysis.
(8) Scan the dataset in order to find the most frequent values of each predictor in cases when the purchase exists. Save the customer profile for this itemset X_i.
(9) Repeat the procedure from Step (3) until all interesting antecedents *X* identified in Phase 1 are used in the output variable.

The above algorithm results in extracted customer profiles for each of the interesting itemset X identified in Phase 1. For example, if an interesting itemset X_i is {*milk, bread*}, then the result of Phase 2 are the characteristics of customer who buy milk and bread together in this store (their age, gender, education level, home ownership, number of kids, number of cars, etc.). The obtained profiles can be used to plan marketing activities for each itemset X_i.

The second component is the **Knowledge management** (KM) aimed to extract knowledge from the results obtained in the previous DM component. It is called **Phase 3** or the "**KNOW**" phase, and contains two main groups of activities: (1) sharing interesting patterns and customers' profiles with employees, and (2) collecting new ideas, ranking, and selecting ideas that have a potential to be transformed into new marketing strategies. The main function of this phase is to

achieve visibility of knowledge extracted from the previous phase to all employees (or to a set of employees involved in knowledge management), and to enable them to actively participate in generating new marketing strategies. Those activities could be technologically supported by a web-based knowledge management system. The user interactivity in this phase includes examining extracted interesting patterns, as well commenting on each pattern and customer's profiles, raising a question, answering a question, brainstorming for generating new ideas for marketing, rating, and ranking the suggested ideas. This phase actually contributes creating and updating a knowledge base of an organization.

The last phase of the model is **Phase 4** or "**HOW**" phase, and it is focused on generating new marketing strategies leaded by marketing managers and sale managers, but also including other employees. In this component, it is important to use the ideas extracted in the KM component, and formulate new strategies, mainly focused on the following types of marketing innovative strategies identified by European Commision (2012): (1) new media or techniques for product promotion, (2) new methods for product placement or sales channels, and (3) new methods of pricing goods or services. The effects of this phase should increase the sale, cross-selling index and competitiveness of the company. They should also serve as a feedback to other model components in order to improve their efficiency. In order to test the model efficiency, an illustrative example of model usage is conducted and described.

Results – Illustrative example

Phase 1 (WHAT) - ARs were used as a DM method suitable for discovering patterns in shopping behaviour, i.e. market basket analysis. After experimenting with different values of support s and confidence c (values ranged from 3% to 50%), the following parameters were selected as best-suited according to the variety of generated rules: minimal support coefficient $s = 10\%$, minimal confidence coefficient $c = 10\%$. The total number of 36 ARs were extracted as significant, while the first 9 rules with confidence c larger than 50% are presented in Table 1. The symbolic representation of rules $X \rightarrow Y$ is used in all tables. ARs that satisfy all three suggested criteria of interestingness (confidence min. 50%, heuristic unexpectedness, and heuristic actionability) are bolded. The notation of items in Table 1 is the following: X_1 = towel, Y_1 = bag, X_2 = tuna, Y_2 = dough, X_3 = cream, Y_3 = toilet paper, X_4 = toilet paper, Y_4 = cheese, X_5 = milk, Y_5 = towel, X_6 = yogurt, Y_6 = yogurt, X_7 = dough, Y_7 = milk, Y_8 = bread.

Table 1. First Nine Association Rules Generated in the WHAT component (min. $s=10\%$, min $c=50\%$)

Rule ID	Rule	Objective measures		Heuristic evaluation of human expert	
		Support (%)	Confidence (%)	Unexpectedness	Actionability
1	$X_1 \rightarrow Y_1$	11,11111	80,00000	0	1
2	**$X_2 \rightarrow Y_2$**	**9,25926**	**71,42857**	**1**	**1**
3	**$X_1 \rightarrow Y_3$**	**9,25926**	**66,66667**	**1**	**1**
4	$X_3 \rightarrow Y_4$	15,74074	60,71429	0	1
5	$X_4 \rightarrow Y_5$	9,25926	58,82353	1	1
6	$X_5 \rightarrow Y_6$	20,37037	51,16279	0	1
7	$X_6 \rightarrow Y_7$	20,37037	51,16279	0	1
8	$X_3 \rightarrow Y_6$	12,96296	50,00000	0	1
9	$X_7 \rightarrow Y_8$	10,18519	50,00000	0	1

Source: Authors

Since it would be difficult to plan marketing strategies based on 20 different patterns, it is necessary to evaluate those patterns by using subjective measures of a human expert. After such heuristic evaluation according to the criteria of heuristic unexpectedness and heuristic actionability, only two rules: $X_2 \rightarrow Y_2$ (*if tuna than dough*) and $X_1 \rightarrow Y_3$ (*if towel than toilet paper*) satisfy all three criteria used in this research (confidence >= 50%, positive unexpectedness, and actionability), therefore those rules can be considered as highly interesting ones. The rule $X_2 \rightarrow Y_2$ is the most interesting one, since it has a very high confidence (71,43%) revealing a high probability that a customer will buy the itemset Y_2 if it buys X_2, and it is also positively evaluated by the expert in terms of heuristic unexpectedness and actionability. The support of this rule is not high (9.26%) revealing that item pairs that have a low frequency could have a high probability and therefore be interesting for decision makers. This rule was also a surprising one to the management of the observed store. The highest confidence (80%) with a support of 11.11% is obtained for the rule $X_1 \rightarrow Y_1$, which was not surprising, and therefore not selected as interesting one. The rule $X_1 \rightarrow Y_3$ also shows a high confidence (66.67%) in spite of a low support (9.26%), and is also unexpected and therefore selected as another interesting rule for further analysis.

The results of the AR method show that the extracted rules were large in quantity, but the quality of rules was not high since there were many rules with a very small support in the dataset. Also, majority of rules were not unexpected and actionable, which is the reason that a very small number of rules were selected as interesting ones after the human expert evaluation.

Phase 2 (WHO) - NN model is created on the basis of the AR results. The 9-step algorithm described in section 3.3 is used to generate the most successful NN model. Since there are two extracted rules in Phase 1 ($X_2 \rightarrow Y_2$ and $X_1 \rightarrow Y_3$), only two NN models will be created in this phase:
  o NN model 1 – with the purchase of X_2 itemset as the output variable P_X
  o NN model 2 – with the purchase of X_1 itemset as the output variable P_X
In both NN models, the set of input variables C consists of 8 available customer characteristics. For the purpose of creating NN models, values for the input variables were software-generated. The total sample was randomly divided into train (70% of data), test (15% of data) and validation subsample (15% of data). Three activation functions were used, and the model which uses tangens hyperbolic activation function is selected as the most efficient one, because it contains a lower number of hidden units. The results of NN model 1 are presented in Table 2.

Table 2
Results of NN Model 1

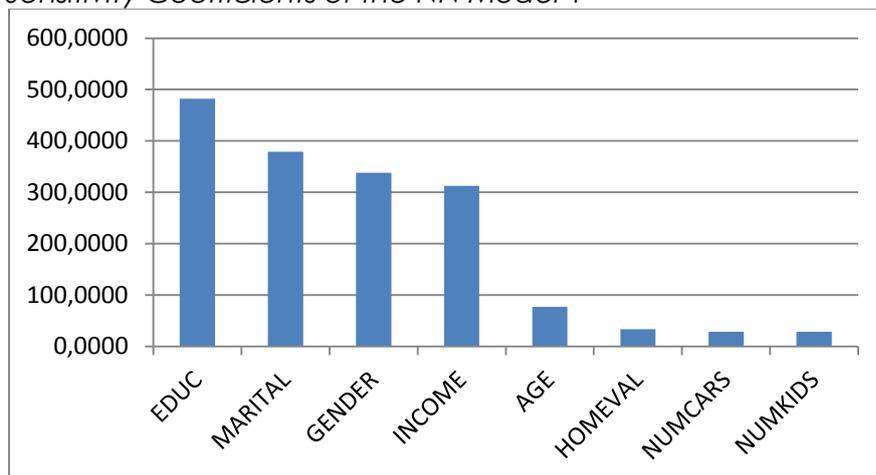| NN Architecture	Activation Function	Classification Accuracy Obtained on the Validation Sample (%)		
		1 (P_X=1)	0 (P_X=0)	Total Classification Rate
18-14-2	Tangens Hyperbolic	100.00	98.59	98.73

Source: Authors

It can be seen in Table 2 that the total classfication rate obtained on the validation sample is 98.73%, meaning that the NN model correctly classifies 98.73% of customers into a correct category of purchase. The NN model is more accurate for the category 1 (which denotes the existance of purchase of itemset *X*), than for the

category 0 (which denotes the absence of purchase of itemset *X*). Considering such a high accuracy, it can be concluded that the NN methodology produces a satisfactory model.

Following the step (7) of the algorithm described in previous section, important predictors of purchase are to be selected by sensitivity analysis. The sensitivity coefficients of selected predictors are presented in Figure 2:

Figure 2
Sensitivity Coefficients of the NN Model 1

Source: Authors

It can be seen that the highest influence on purchasing itemset X_2 is obtained by the *education level, then marital status, gender,* and *income*, while other variables, such as *age, home values, number of cars,* and *number of kids* have significantly lower influence to the model accuracy, although their influence is positive. The next step of the algorithm for discovering the customers' profile is to scan the dataset in order to find the most frequent values of each predictor in cases when the purchase exists.

In this a-posterior analysis of each input predictor with the selection of only category of $P_X=1$, the following customers' profile for customers of who purchased the itemset *X* is obtained:

o *Gender="female"* for 64% of customers who purchased the itemset *X*.
o The most frequent category of *Education level is "5" (college of university level)* (39.87%).
o The most frequent category of *Marital status="single"* (34.17%).
o The most frequent category of *Income level = "middle"* (44.63%).
o Number of kids = 0 (mode value).
o *Number of cars =* 1 (mode value).
o *Age* = 30 (mode value).
o *Home value* = 173200 *HRK* (mode value).

Therefore, the typical customer who most frequently buys itemset X_2 is a single female aged 30, has a college or university degree, a middle level income, has one car, no children, and its home value is about 173200 HRK. The procedure of identifying customers' profile can be repeated for all interesting antecedents *X* identified in Phase 1 (i.e. for the itemset X_1).The marketing managers could use the extracted customers' profiles in generating new ideas about their marketing strategies in the KM system, which is the next component of the DataKnow model.

Phase 3 (KNOW) - The activities in the KM component: (1) sharing interesting patterns with employees, and (2) generating new ideas, ranking, and selecting ideas, can be implemented by a KM software tool. A number of such tools are available today, and the choice should be dependent of the functionalities needed and the affordable cost. For the purpose of testing the proposed model, the list of features should include the following abilities: to add an article (i.e. document) and workspace, to share the articles and workspaces, to collaborate, to comment an article, to pose questions, to answer questions, to rate (or rank) documents and comments, and brainstorming. Additional functionalities that are desirable are the ability to create blogs, wikis, and share applications. Such knowledge sharing is able to contribute organizational learning and generate some new ideas on marketing strategies that could be made upon extracted unusual patterns.

Phase 4 (HOW) - The new marketing strategies that could be generated in the last, HOW component of the model include for example the following:
o   new techniques for promoting itemset X_i by enabling a customer to virtually use it and give some extra advices on usage together with itemset Y_j
o   new ways to place itemset X_i in relation to itemset Y_j on a shell or in a web or mobile store application,
o   new methods of pricing goods or services of itemset X_i together with itemset Y_j
   The new marketing strategies could be based on web 2.0 and web 3.0 concepts that enable customer interaction, individual custom design of products or promotions, etc.

Discussion

The results show a high potential of the suggested integrative DataKnow model in generating new marketing strategies. The suggested DataKnow model consists of (1) DM component for discovering patterns of shopping behavior and customer profiles, and (2) KM component for sharing patterns and extracting knowledge that will assist in generating new ideas for marketing strategies. In the DM component, ARs generated two interesting patterns of shopping behavior which were used in the NN model to identify profiles of customers that buy those frequent items. The experimental research showed that the combination of objective and subjective measures in extracting ARs could be efficient way to include marketing managers as active participants of a knowledge managements system. A high total accuracy of the NN model (97.73%) in recognizing customers who are likely to purchase a frequent itemsets is very promising. The NN methodology is able to incorporate extracted rules and give deeper insight into customer profiles. The suggested model enables all involved employees to use the extracted patterns and customer profiles as triggers for new ideas in defining creative marketing strategies in the KM component of the model.

   If implemented by the management, the model could result with a higher cross-selling index and higher customer satisfaction. The model could influence marketing by enabling systematic support for generating new knowledge from the data about customer purchases through organizational learning by revealing unexpected patterns in customer behavior.

Conclusion

The aim of the paper was to suggest a model that will integrate data mining based on association rules and neural networks in knowledge management such that it

generates new marketing strategies. Association rules were used as a data mining method to discover which products are frequently purchased together or sequentially by the same customer. In order to extract interesting rules that are confident, but also unexpected and actionable to the user, an integration of objective measures and heuristic subjective evaluation of human expert is proposed and tested. The results showed that such procedure is able to generate few interesting rules. The itemsets extracted in the most interesting rules are used in the second phase to create the output variable of the neural network model. The results showed that the NN model successfully finds which customer characteristics are important predictors of a purchase, and is able to reveal the profiles of customers who are likely to buy an itemset. The extracted association rules, together with customer profiles can be used in a knowledge management (KM) system to generate new strategies in marketing.

Since this is a preliminary research, the paper has some limitations, such as a single dataset used to illustrate the model efficiency. The future studies should focus on including more datasets in order to achieve model generalization, and on testing additional data mining methods, such as support vector machines and others. Further limitations of the modelling procedure are in selection of experts and their human judgement in extracting association rules, which can be time-consuming and subject to mistakes, therefore should be dealt in future research. Also, other dimensions of the knowledge management component should be analyzed that are not based only upon customer profiles and shopping patterns.

The research could be useful to marketing and retail managers to improve the process of their decision making in order to generate more innovative marketing and therefore competitive advantage of a company, as well to researchers in the area of marketing modelling.

References

1. Agrawal, R., Imielinski, T., & Swami, A. (1993), "Database mining: a performance perspective", IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 6, pp. 914-925.
2. Agrawal, R., Srikant, R. (1994). "Fast Algorithms for Mining Association Rules", in Bocca, J. B., Jarke, M., Zaniolo, C. (Eds.), Proceedings of 20th International Conference on Very Large Data Bases VLDB '94, September 12-15, 1994, Morgan Kaufmann, Santiago de Chile, Chile, pp. 487-499.
3. Chen, C. H., Lan, G. C., Hong, T. P., Lin, Y. K. (2013), "Mining high coherent association rules with consideration of support measure", Expert Systems with Applications, Vol. 40, No. 16, pp. 6531–6537.
4. European Commission (2012), "EuroStat, The Community Innovation Survey 2012", available at http://ec.europa.eu/eurostat/documents/203647/203701/Harmonised+survey+questionnaire+2012/164dfdfd-7f97-4b98-b7b5-80d4e32e73ee (15 April 2015).
5. Heinrichs, J. H., Lim J. S. (2003), "Integrating web-based data mining tools with business models for knowledge management", Decision Support Systems, Vol. 35, No. 1, pp. 103-112.
6. Javaheri, S. F., Sepehri, M. M., Teimourpour, B. (2013), "Response Modeling in Direct Marketing: A Data Mining-Based Approach for Target Selection", in Zhao, Y., Cen J. (Eds.), Data Mining Applications with R, Amsterdam, Elsevier, pp. 153–180.
7. Khan, D. M., Mohamudally, N., Babajee, D. K. R. (2013), "A Unified Theoretical Framework for Data Mining", available at

http://dx.doi.org/10.1016/j.procs.2013.05.015 (15 April 2015).

8. Kotler, P., Armstrong, G. (2010), Principles of marketing, Pearson Education.
9. Liao, C. W., Perng, Y. H., Chiang, T. L. (2009), "Discovery of unapparent association rules based on extracted probability", Decision Support Systems, Vol. 47, No. 4, pp. 354–363.
10. Lin, K. C., Liao, I. E., Chen, Z. S. (2011), "An improved frequent pattern growth method for mining association rules", Expert Systems with Applications, Vol. 38, No. 5, pp. 5154-5161.
11. Liu, B., Ma, Y., Wong, C. K. (2001), "Classification Using Association Rules: Weaknesses and Enhancements", in Grossman R. L. et al. (Eds.), Data Mining for Scientific and Engineering Applications, Springer, available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.8.7943&rep=rep1&type=pdf (19 July 2014).
12. Masters, T. (1995), Advanced Algorithms for Neural Networks, A C++ Sourcebook, New York, USA, John Wiley & Sons, Inc.
13. Paliwal, M., Kumar, U. A. (2009), "Neural networks and statistical techniques: A review of applications", Expert Systems with Applications, Vol. 36, pp. 2–17.
14. Silberschatz, A., Tuzhilin, A. (1995), "On subjective measures of interestingness in knowledge discovery", in Fayyad, U. M., Uthurusamy, R. (Eds.), Proceedings from the First International Conference on Knowledge Discovery and Data mining (KDD-95), August 20-21, 1995, Montreal, Canada, pp. 275-281.
15. Rygielski, C., Wang, J. C., Yen, D. C. (2002), "Data mining techniques for customer relationship management", Technology in Society, Vol. 24, No. 4, pp. 483-502.
16. Shaw, M. J., Subramaniam, C., Tan, G. W., Welge M. E. (2001), "Knowledge management and data mining for marketing", Decision Support Systems, Vol. 31, No. 1, pp. 127-137.
17. Tsai, H. H. (2013), "Knowledge management vs. data mining: Research trend, forecast and citation approach", Expert Systems with Applications, Vol. 40, No. 8, pp. 3160–3173.

About the authors

Marijana Zekić-Sušac is a full professor at the University of J.J. Strossmayer in Osijek, Faculty of Economics in Osijek, Croatia. She has earned her doctoral degree at University of Zagreb, Faculty of Organization and Informatics Varaždin, Croatia. Her research interests include artificial intelligence, machine learning and data mining in business, education and medicine. She currently teaches several ICT courses on undergraduate, graduate and doctoral level. She is a member of the International Neural Network Society and the president of the Croatian Operational Research Society. Author can be contacted at **marijana@efos.hr**

Adela Has was graduated in 2010 at Faculty of Economics in Osijek, University of J.J. Strossmayer in Osijek. She is a doctoral student at International inter-university postgraduate interdisciplinary doctoral program Entrepreneurship and Innovativeness. She is employed as an assistant at the Faculty of Economics in Osijek for the scientific field of economics, business information branch. She is a member of the Croatian Operational Research Society. Author can be contacted at **adela.has@efos.hr**