Association Rules as a Decision Making Model in the Textile Industry

DOI: 10.5604/01.3001.0010.2302

Abstract
Sales process disfunctions in the textile industry are problems that cause loss of customers, incomplete market supply, etc. The objective of the research is to analyse transactions from the textile industry database in order to find patterns in buyers’ behavior and improve the model of decision-making. Association rules, one of the most noticeable data mining techniques, is used as methodology to learn rules and market patterns that occur in sales in the textile industry, which will enhance the decision-making process, by making it more effective and efficient. The Apriori algorithm was applied and open source software Orange was used. It has been shown using a real-life dataset containing 2000 transactions from the textile industry of the South East Europe region that the approach proposed is useful in discovering effective knowledge in data associated with sales. The study reports new interesting rules and the dependence of the following parameters: support, confidence, lift and leverage on making more customized offers in the textile industry.

Key words: business intelligence, association rules, decision making, textile industry.

Introduction
The application of data mining and knowledge discovery has attracted both researchers and industry because of its ability to find new and useful patterns in data in order to acquire knowledge and apply it in practice. One of the most important tendencies of modern business is the possibility to create a more customized offer of products and services [1], which is present in the textile industry. By creating customized offers, the company would benefit from more satisfied customers, leading to a bigger market share and income. In this paper, we tackle the following problem. The sales process in the textile industry needs customisation in order to fulfill the needs of individual customers. However, it is really hard to identify features explaining a phenomenon which could easily lead to poor recommendation, most often being the mass confusion problem [2]. The development of appropriate recommender systems in the textile industry which can be used as a decision support system can tackle the mass confusion problem. This problem is interesting to both research and the industry community.

In order to address this problem, our work advances the current state of the art by utilising association rules based essentially on customer needs. Our association rule model takes into account the analysis through data mining of a national retailer (including part of the textile industry) that was explained in [3]. The advantage of the application of association rules in the textile industry is the ability to reduce searching time and cost to find patterns available in data that should not be neglected. This is further used as a recommender system which may recommend a product to the user that should match the preferences. However, a poorly developed system can recommend a product which matches a small subpopulation, leading to defective recommendation. Therefore, how to apply data mining methods and tools to assist association rule mining has become an interesting research issue. In addition, the number of textile brands is growing fast, which is interesting to the research community (how to find association among products in the big data era). To sum up, this study aims to analyse sales in the textile industry using association rules to help decision-making by establishing a recommendation system. In the following section, we review the state of the art for association rules, as well as the decision support system in the textile industry. Furthermore, we describe methodology used in this paper, followed by the experimental design and results. Finally, findings are discussed and benefits of this approach are concluded alongside future work.

Theoretical background
In this section we present the state of the art in areas that are associated with our work in a complementary way. Furthermore, we discuss the differences between our work and the current state of the art. The application of business intelligence in the textile industry is not much explored yet. According to [29], Agrawal was first to establish association rules back in 1993 in order to
perform an analysis of a market basket. The market basket method [32-34] has predominantly been used in sales analyses since, although its significance has been proved in other fields: the analysis of credit card sales [35, 36], identification of insurance companies committing fraud, or the analysis of telecommunication services [5]. However, this method cannot be explicitly used in the case of simultaneous events, but only in the case of successive events, which may be very useful in marketing for instance [37, 38]. All the examples hereabove show how wide the application of association rule methods in business can be. The authors have drawn some significant conclusions based on what right business decision can be taken.

To the best of our knowledge, the application of association rule mining in the textile industry is seldom used. In [39] the mining of association rules in the garment industry has been explained. The authors believe that association rules (consequence of the rule is the same in all rules) could be more efficiently solved by using classification algorithms, such as logistic regression. In our paper, we employ association rules to find any relation between any two or more features in our data. This approach will yield more interesting rules in the textile industry. The application of the multi-criteria decision-making technique AHP in the textile industry has been described in [43]. This paper presents an interesting approach to select a suitable ERP (Enterprise resource planning) system for the textile industry. The advantages of the business intelligence tool – AHP in the textile industry have been well described. Organisations tend to invest in business intelligence solutions in order to support the decision-making process. An interesting study on free business intelligence system implementation has been described in [44]. There is also an original approach that presents customer segmentation in a large database of an online customised fashion business. The authors solve marketing and manufacturing problems by the use of data mining in the fashion industry [45]. The goal of this paper was to investigate two different data mining approaches for customer segmentation: clustering and subgroup discovery. The models obtained produced six market segments and forty-nine rules which provided a better understanding of customer preferences in a highly customised fashion business.

Some of the recent trends and developments of textile business and intelligent systems that are breaking the bounds of traditional textiles and their design have been described in [46]. There is active research related to intelligent clothing where applications combine electronics and IT with textiles. Different perspectives on the textile and clothing industry in Greece have been described in [47]. Innovations adopted in the production process, new cost-efficient technologies and workers’ training are the keystones in establishing a competitive industry for the future [47]. IT also has an important role in creating efficient textile industry innovations. The authors conducted research among companies of the textile and clothing industry to compare the predictive ability of five developed models based on three statistical techniques (discriminant analysis, logit and probit) and two models based on artificial intelligence (neural networks and rough sets). This research could be a great contribution to devisers of national economic policies that aim to reduce industrial unemployment [48]. This shows the great potential and importance of artificial intelligence applied in the textile and clothing industry. There is research on recommender systems that have influenced environments where data size exceeds the capabilities of any user to fully explore the available choices in the store (physical or on-line). Authors use association rules to provide recommendations to customers, as well as to understand who the customer is and what their needs are when entering a physical store or the corresponding e-shop [49]. This research shows that the rules generated, expressing the customers’ shopping behavior, can feed a recommender system that proposes new garments to customers based on these rules.

### Methodology

#### Association rules

Data mining is a process of analysing large data sets in order to discover significant patterns and rules. As modern companies are constantly seeking higher goals regarding productivity, it becomes absolutely necessary to improve the functioning of their organization through better understanding of their customers’ needs. Data mining techniques and tools are widely used in various fields of application – law, astronomy, medicine, industrial process control, etc [13, 15, 21]. With the rapid development of information technologies, data are becoming larger and larger, demanding new approaches in data mining [16]. Data mining is mostly used for creating models [9]. According to [12], association rules were first introduced by Agrawal. An association rule is an expression $X \Rightarrow Y$, where $X$ and $Y$ are sets of items. This means that the transactions of the database which contain $X$ tend to contain $Y$ [12]. The task is to identify a set of rules which co-exist in some data set. There is a case of association rules in the field of market basket [10], where these rules classify into groups the items purchased together at supermarkets. Data mining algorithms can be applied to the Internet of Things by extracting hidden information from raw data [19]. An interesting example of the assessment of data quality in accounting with association rules has been proposed [25]. The supplier selection problem can be interpreted as a multi-criteria decision-making problem and therefore requires complex data mining tools. An example of the use of the analytic network process and data envelopment analysis (DEA) has been described in [27].

The quality of association rules discovered, and therefore their importance, is estimated based on two parameters – support and confidence. These parameters are defined in [16]:

\[
\text{Confidence} \ (X \Rightarrow Y) = \frac{P(Y | X)}{P(X)} \quad (1)
\]

\[
\text{Support} \ (X \Rightarrow Y) = P(X U Y) \quad (2)
\]

X and Y represent sets of items.

Confidence is defined as the part of a sample to which a certain association rule applies [29]. It is given in % and describes how well this rule is established in the sample. If we now state that each sample consists of a number of cases, then the confidence determines what percentage of cases possessing attribute X also possess attribute Y. Support has been defined as the number of transactions that contain an itemset [31]. A very challenging issue of the process of association rule discovery is finding rules for attributes with numerical values [4]. The greatest disadvantage of association rules is the lack of sufficient interconnectivity between them in most cases when large databases are analysed, in which case their simplicity may lead to their inconsistency [8]. The apriori algorithm distinguishes itself as the most basic one, and is based on the presumption that if a number of items appears in a data set,
in which case each item appears in the same data set as well [11]. The benefits of the improved Apriori algorithm have been elaborated [6, 17]. A multilevel association rule algorithm has been presented [24]. Other algorithms available for association rule mining have been compared and analysed [7].

A model of the Apriori algorithm has been presented [42] in Figure 1.

According to [42], the Apriori algorithm counts the item occurrences to determine the large I – sets of items. Then the subsequent pass k consists of two phases: first the large sets of items are found in the (k-1) pass are used to generate candidate sets of items using the Apriori function. Then the database is scanned and the support of candidates of is counted. Efficient determination of candidates in that are contained in a given transaction t needs to be done. Mass customisation is a production strategy focused on the broad provision of personalised products and services, mostly through modularised product/service design, flexible processes and integration between supply chain members [26]. An improved Apriori algorithm has been described with application to a textile dataset [22, 28]. The use of business intelligence (AHP) in the garment industry has been described in [40]. An interesting study on developing a recommender system using association rules with the Apriori algorithm for data analysis has been presented in [41].

In order to discover association rules in a large database of sales transactions in the textile industry, data are processed using Orange software [14], which is a very popular data mining tool among managers, who use it widely to support themselves in decision-making. Orange is a comprehensive, component-based software suite for machine learning and data mining. Orange performs numerous data mining tasks, such as clustering, classification, association rules, etc.

**Database**

Data used for the purpose of this research are all related to textile industry transactions. A total of 2000 transactions were processed from data of real business systems of the textile industry from the South-East Europe region during a six month period. Data attributes were assigned to six categories: buying rates (very high, high, medium and low rate of transactions), fibre maintenance (very high, high, medium, low level), number of texture (two, three, four or five), number of buyers (two, three, four), fibre size (small, medium, large) and fibre quality (low, medium, high). Data are described in a qualitative manner with four adjectives: accurate, inaccurate, good and very good. In Figure 2 there is an overview of the variable buying rate, with its clear distribution of all four categories (low, medium, high and very high). The highest buying rate is represented by the inaccurate class (blue color), whereas the red one is for the accurate class. Orange (textile class: very good) and green (textile class: good) colors show poor buying rates.

**Experiment**

The main hypothesis of the research is proving that association rules have a significantly important application in the textile industry, as well as creating patterns from databases that could lead to efficient decision-making. An experiment was set up in real-term business settings. The sub-hypothesis of the experiment is showing whether results of the association rules obtained in the textile industry are dependent on the definition of support and confidence parameters and in what manner.

The objectives of the paper are:
- The analysis of transactions from the textile industry database for a period of six months in order to define an improvement of the textile industry sales model;
- The application of a business intelligence tool (association rules) by means of open source software Orange in order to find patterns important for decision-making in the textile industry;
- Development of a model of association rules which shows the change

---

**Figure 1.** Apriori algorithm.

**Figure 2.** Overview of textile buying variable.

**Figure 3.** Overview of the dependence of association rules and change in parameters.
of parameter support, confidence, lift and leverage;
- Calculation of the standard deviation of parameter support, confidence, lift and leverage;
- Presentation of the model of association rules achieved with recommendations of buyers’ behaviour in the textile industry.

According to [30], the three aspects that determine the complete development of modern decision making are qualitativity, quantitativity and information-communication. The results of the experiment are shown in respect of all three aspects of business decision-making. Data are processed into numerical expressions to ensure the quantitative aspect of modern decision-making. Results are analysed in order to confirm or not the main hypothesis of the research. The association rules obtained help managers in decision-making by providing them with relevant data on customers’ buying habits. Validation of both the project solution obtained and its application in practice has been made as well as. As rules selection criteria we set support to be minimum 15%, and confidence minimum 30%. Consequently we increased support to 20% and confidence to minimum 40%. Finally we increased support to be minimum 30% and confidence to be minimum 50%. Results were recorded and analysed.

Change in the predefined parameters resulted in a change in the number of association rules obtained. Figure 3 clearly shows the dependence between the number of association rules obtained and the change in predefined parameters of support and confidence (we analysed three predefined changes). Consequently the bigger values of support and confidence parameters resulted in a smaller number of association rules.

We increased the support to minimum 20% and confidence to minimum 40%.

**Figure 4.** Overview of parameters for 25 association rules.

**Figure 5.** Overview of parameters for 10 association rules.

**Figure 6.** Overview of parameters for 2 association rules.

**Figure 7.** Overview of values of standard deviation of all four parameters.

**Figures 4, 5 and 6** show all values of the four parameters (support, confidence, lift and leverage) and their change. We can note that Figure 4, with the largest number of association rules, also has the highest line peaks. Figure 5 shows a smaller number of association rules, which result in even more lines (values) of the parameters (with small peaks). In Figure 6 there is an example of two association rules with two equal values. As rules selection criteria we set support to be minimum 15% and confidence minimum 30%.

Figure 4 shows that rules number fifteen and nineteen have the biggest line peaks (values of parameters). Therefore there is a strong likelihood that these two association rules will show up in the database. If a customer purchases a two-fibre textile (rule No. 15), they will most likely decide to purchase the one classified as "inaccurate". If a customer purchases low-quality textile (rule No. 19), they will most likely decide to purchase the one classified as "inaccurate".
**Table 1. Standard deviation of support, confidence, lift and leverage.**

<table>
<thead>
<tr>
<th>Number of association rules</th>
<th>ST DEV support</th>
<th>ST DEV confidence</th>
<th>ST DEV lift</th>
<th>ST DEV leverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.059287</td>
<td>0.179691606</td>
<td>0.210734</td>
<td>0.047970894</td>
</tr>
<tr>
<td>10</td>
<td>0.060119</td>
<td>0.18993882</td>
<td>0.223431</td>
<td>0.052507</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 2. Ten association rules.**

<table>
<thead>
<tr>
<th>Supp</th>
<th>Conf</th>
<th>Lift</th>
<th>Leverage</th>
<th>Antecedent</th>
<th>Consequent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.333</td>
<td>0.476</td>
<td>1.428</td>
<td>0.100</td>
<td>y = inaccurate</td>
<td>buyers = 2</td>
</tr>
<tr>
<td>0.333</td>
<td>0.476</td>
<td>1.428</td>
<td>0.100</td>
<td>y = inaccurate</td>
<td>quality = low</td>
</tr>
<tr>
<td>0.207</td>
<td>0.620</td>
<td>0.885</td>
<td>-0.027</td>
<td>quality = medium</td>
<td>y = inaccurate</td>
</tr>
<tr>
<td>0.333</td>
<td>1.000</td>
<td>1.428</td>
<td>0.100</td>
<td>y = inaccurate</td>
<td>quality = low</td>
</tr>
<tr>
<td>0.333</td>
<td>1.000</td>
<td>1.428</td>
<td>0.100</td>
<td>y = inaccurate</td>
<td>buyers = 2</td>
</tr>
<tr>
<td>0.208</td>
<td>0.833</td>
<td>1.190</td>
<td>0.033</td>
<td>maintenance = very-high</td>
<td>y = inaccurate</td>
</tr>
<tr>
<td>0.260</td>
<td>0.781</td>
<td>1.116</td>
<td>0.027</td>
<td>size = small</td>
<td>y = inaccurate</td>
</tr>
<tr>
<td>0.227</td>
<td>0.681</td>
<td>0.972</td>
<td>-0.007</td>
<td>size = medium</td>
<td>y = inaccurate</td>
</tr>
<tr>
<td>0.213</td>
<td>0.639</td>
<td>0.912</td>
<td>-0.020</td>
<td>size = big</td>
<td>y = inaccurate</td>
</tr>
<tr>
<td>0.208</td>
<td>0.833</td>
<td>1.190</td>
<td>0.033</td>
<td>buying = very-high</td>
<td>y = inaccurate</td>
</tr>
</tbody>
</table>

**Figure 5** shows an overview of parameter values for the ten association rules obtained. Of all the other rules, we can note that No. 3 has the biggest decline parameter values (lift, support and leverage). Lift has the biggest decline, according to rule No. 3. We can also note that rules No. 4 and 5 have the biggest values of all four parameters, and are therefore distinguished as the strongest ones. There is a strong likelihood that rules No. 4 and 5 will show up in the database. Finally we increased support to be minimum 30% and confidence to be minimum 50%. **Figure 6** shows the two association rules obtained with the same values of all four parameters. There is the same probability that these two rules show up. We conclude that the higher values of predefined parameters (support and confidence) cause a lower number of association rules obtained.

**Table 1** shows an overview of the standard deviation of parameters support, confidence, lift and leverage for all three predefined parameters’ change. We can note that the third parameter’s change resulting in two association rules shows the most precise measurement (standard deviation is zero). Then we can also note that the parameter leverage in the second predefined change that resulted in ten association rules is very precise (standard deviation is 0.052507), when compared to other the parameters’ values.

In **Figure 7** there is graphical overview of standard deviation values of the parameters. We can see that the biggest value of standard deviation is shown by the lift parameter (regarding the 25 and 10 rules). This means that this parameter is the least precise. The standard deviation of the parameter leverage has the lowest value (regarding the 25 and 10 rules). The standard deviation of all four parameters (regarding two rules) is zero, because all results in the distribution are equal. The lower the standard deviation, the more precise the method used.

**Results and discussion**

The association rules obtained are presented in **Table 2**.

Support = 20%
Confidence = 40%
Number of association rules: 10

There are two highlighted association rules with the highest values of parameters support, confidence, lift and leverage: 0.333, 1.000, 1.428 and 0.100, respectively. First the highlighted rule is that if a customer purchases a low-quality textile then he will most likely decide to purchase the one classified as “inaccurate”. The second highlighted rule is that if there are two buyers, they will most likely buy from the class “inaccurate”. The analysis of each association rule shows precision in characteristics and patterns in customers’ behaviour. Different results were obtained when support was changed from 15 to 20% and confidence increased from 30 to 40%. The number of association rules discovered consequently decreased from 25 to 10. There were ten new association rules with support ranging from 20 to 33% and confidence from 47 to 100%. These association rules create very interesting and significant patterns in customer behaviour in the textile industry. Then the knowledge created from databases could lead to efficient decision-making that results in increased profit in the textile industry.

Furthermore considering the fact that support was set at a relatively high 30%, only two association rules were obtained. If quality is low, then fibres will belong to the class “inaccurate”. If there are two buyers, they will most likely buy from the class “inaccurate”. These rules can be used to put all fibres in shops with low quality into the class “inaccurate”. If two buyers come together, then a customised offer with a special promotion could be offered to them from the fibre class “inaccurate” because there is the strongest possibility to buy it. The Apriori algorithm was applied to data and a set of association rules was obtained [18]. Rules were then checked and the most important ones selected for creating meta rules, in order to reduce redundancy and obtain the most acceptable solution.

Issues of environmental protection and waste management in the textile industry have been extremely important. Real-time settings and practical implication of research provide advantage when creating an efficient textile industry ecosystem. The multidisciplinary approach of the research provides the dependance between business intelligence, human resource management, the textile industry and modern ICT. In creating effective business system management with a multiple cost-saving impact on the environment. The problem discussed in the paper results in the recommendation of association rules gained that could be used as guidance for analysing buyers’ behaviour. Then the sales department can organise better sales facilities with a customized offer in order to increase the buying rates and profit. Results show that higher values of parameters (support, confidence, lift and leverage) result in a smaller number of association rules. This research result relates to environmental protection with a positive impact because of the precise creation of knowledge from a database and better resource allocation customized according to buyers’ behaviour.
One of the directions for future research could be the use of association rules for creating an improved textile industry sales model for better organisation and distribution of a textile supply chain and its resources as well as a more efficient waste management system in the textile industry. In addition, future research could be to create a textile industry sales model with association rules that shows the dependence and impact of predefined parameters (e.g. absorbency, chemical resistance, flammability, strength of fibres) on environmental protection. The main advantage of this research is its original approach in using business intelligence tools based on science [20]. This new model based on the association rules obtained is developed to solve business problems efficiently. The research topic is extremely challenging and provides extensive space for further elaboration of association rule application. The association rules obtained were used to help in creating comprehensive knowledge on certain aspect of decision-making in the textile industry [23]. Experienced business intelligence analysts were consulted regarding research results.

## Conclusion

By analyzing current research work of various experts in the field of business intelligence, and despite the fact that this field is relatively new in some segments of application, the great popularity and potential have been proved. Therefore it presents a challenge to researchers in those areas where a further significant scientific and expert contribution can be expected. Any innovative research in the field of business intelligence is shaped with knowledge and creativity and is conducted with the assistance of modern data mining software architectures. The results obtained in the textile industry have been analyzed using modern scientific methods and several recommendations for a future research course have been made. Our research confirmed the hypothesis that association rules have important application in the textile industry, while creating patterns from databases that could lead to efficient decision-making. The experiment confirmed the sub-hypothesis of the paper that the number of association rules obtained is dependent on the definition of the support and confidence parameters. The bigger values of predefined parameters support/confidence create a smaller number of association rules in textile industry application. The process of improving the textile industry decision-making model has been presented taking into consideration that managers learn about their customers’ purchase habits through association rules. The process of decision-making has been improved because of extracting knowledge through patterns from the database of the textile industry by the business intelligence technique. The results of this successfully conducted research are reckoned to be a higher sales rate and better satisfaction of customers. The impact on the environment is multiple cost-saving. Recommended future research in this field is creating a commercial decision-making model applicable to markets belonging to different business systems. The significance of the multidisciplinary approach has been explained, and the dependence between decision making, business intelligence, human resource management, the textile industry and modern ICT has been proven to be effective in the case of business system management.

## References

29. Suknović M, Delibašić B. Business intelligence and systems for support of decision making, FOS, Belgrade. 2010.