A Dynamic Application of Market Basket Analysis with R and Shiny in The Electric Materials Sector[1]

Araştırma Makalesi/Research Article

Gökçe KARAHAN ADALI[1], M. Erdal BALABAN[2]

[1]Department of Business Informatics, Haliç University, İstanbul, Turkey
[2]Department of Management Information Systems, Gelişim University, İstanbul, Turkey
gokceadali@halic.edu.tr, mebalaban@gelisim.edu.tr
(Geliş/Received:26.07.2018; Kabul/Accepted:28.03.2019)
DOI: 10.17671/gazibtd.448245

***Abstract*—** In this study, it is aimed to determine the products that customers prefer to buy together by using algorithm of association rules Apriori, and to implement an application related to customer relationship management. The data set used in this study was obtained from a company operating in the electricity sector. CRoss-Industry Standard Process for Data Mining (CRISP-DM) model was used during data analysis. Association rules technique Apriori applied to data that contains the two year. Data analysis is performed with R language. RStudio was used as a development tool for R codes. The model performed with Apriori was transferred to web environment via Shiny (shinyapps.io). The cross sales realized between HESNYA-03 and HESNYA-02 products of this brand and the sales of different colors together bring the product into the forefront. From this result, it is thought that by creating various color packages related to the product, customers can increase purchases of less preferred products by being offered new products besides these and similar packages. The user is given the opportunity to query the analyzed data set and make basic arrangements related to the algorithm. This allows the application to be dynamic, independent of time and space.

***Keywords*—** *data mining, association rules, market basket analysis, apriori, customer relationship management*

Elektrik Malzemeleri Sektöründe R ve Shiny ile Pazar Sepet Analizine Yönelik Dinamik Bir Uygulama

***Özet*—** Bu çalışmada, birliktelik kurallarına ait algoritmalar kullanılarak, müşterilerin birlikte satın almayı tercih ettikleri ürünler tespit edilerek, müşteri ilişkileri yönetimine ilişkin bir uygulama gerçekleştirilmesi amaçlanmıştır. Çalışmada kullanılan veri seti elektrik sektöründe faaliyet gösteren bir firmadan temin edilmiştir. Veri analizi süresince Çarpraz Endüstri Standard Süreç Modeli (CRISP-DM) modeli takip edilmiştir. İki yılı kapsayan veriye, birliktelik kurallarından Apriori Algoritması uygulanmıştır. Veri analizleri R programlama dili ile gerçekleştirilmiştir. Kodların gerçekleştirilmesinde RStudio geliştirme ortamından yararlanılmıştır. Apriori Algoritması'ndan elde edilen model Shiny (shiny.apps.io) aracılığı ile web ortamına taşınmıştır. HESNYA-03 ve HESNYA-02 ürünleri arasındaki çapraz satış ve ürünün farklı renklerinin birlikte satışı, ürünü ön plana çıkarmaktadır. Buradan hareketle, çeşitli renk paletleri oluşturarak, müşterilerin bu ve benzeri paketlerin yanında yeni ürünler sunarak daha az tercih edilen ürünlerin alımlarını arttırabileceği ortaya konuşmuştur. Kullanıcıya analiz edilen veri setini sorgulama ve algoritma ile ilgili temel düzenlemeleri yapabilme imkânı verilmiştir. Böylelikle uygulamanın zaman ve mekândan bağımsız, dinamik bir hal alması sağlanmıştır.

Anahtar Kelimeler*— *veri madenciliği, birliktelik kuralları, pazar sepet analizi, apriori, müşteri ilişkileri yönetimi

[1] Bu çalışma; Gökçe KARAHAN ADALI tarafından Prof. Dr. M. Erdal BALABAN danışmanlığında hazırlanan "Veri Madenciliğinde Birliktelik Yöntemleri Ve Müşteri Ilişkileri Yönetimine İlişkin Bir Uygulama" başlıklı doktora tezinden derlenmiştir. İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Şubat 2017.

1. INTRODUCTION

In the information age we are going through, it is important that data become meaningful and available knowledge by passing through various data processes. Increasing amounts of data, increasing access channels to data, and also increasing data access speed at the same rate have brought about the emergence of big data concepts. When big data is handled correctly and analyzed through the right information systems, it plays a major role in helping companies at the stage of taking strategic decisions. Customer Relationship Management (CRM) strategies, which have initiated product customization according to the customers that companies have developed while they have been trying to maintain their satisfied customer potential, live on big data. Data mining is needed to create customer understanding that is crucial for building CRM strategy. Determining customers' purchasing preferences, creating customer specific product options, trying to keep the existing customers and getting new customer potential play a very important role in increasing profit of the company.

One of the ways of selling more products to customer is to determine which products or services have a tendency to be bought by customer during a shopping or in successive purchases [1]. Association rules are widely used in data mining, most commonly in the name of market basket analysis [2]. Market basket analysis examines the combination of products purchased by the customer during a shopping experience. These models can be used for market basket analysis and for bundles of products or services that can be sold together [3].

This process analyzes customer buying habits by finding associations between different items that customers place in their shopping baskets. The discovery of such associations can help retailers develop marketing strategies by giving them an idea of which items are often purchased together by customers. The basket analysis also shows the distribution of customers at different perspective. This distribution helps in the process of taking decisions in the form of information planning, advertising design, discount-promotion, store layout and product investment [4]. For instance, if customers are buying milk, how likely are they to also buy bread on the same trip to the supermarket? Such information can lead to increased sales by helping retailers do selective marketing and plan their shelf space [5]. Such association-relationship patterns are potentially present only when the items in the pattern are repeated in more than one transaction [6].

Association rules, as one of the data mining methods used to identify the purchasing trends, have fields of application in many areas of trade, finance, engineering, science and health sectors [2]. They can be used in the analysis of business sales, cross-sell programs, market basket analysis, catalog and placement layout designs and promotional analysis, also in text search operations as well as in the analysis of the frequency of which website pages are clicked together by the same visitors [7].

In this study, it is aimed to create a guiding application within the scope of customer relationship management, through sales data of a company operating in the electricity sector covering the years 2014-2015, in order to better understand the company's customers, to determine their purchasing behaviors, and to develop campaigns, marketing and sales strategies in this direction by separating them into segments. The association rule model (by applying Apriori Algorithm) obtained in the application has been moved to the web environment so that is made accessible and applicable by all. By the developed model, the dynamic applicability of possible product packages on a customer or customer groups in line with the user preferences has been enabled. With this aspect of the study, it is intended to illuminate the studies to be conducted in the similar fields.

2. LITERATURE REVIEW

Frequent itemset mining leads to the discovery of associations and correlations among items in large transactional or relational data sets. With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining such patterns from their databases. The discovery of interesting correlation relationships among huge amounts of business transaction records can help in many business decision-making processes, such as catalog design, cross-marketing, and customer shopping behavior analysis [8].

When the literature is examined, it is seen that association rules can be used not only in the field of marketing, but also in analyzing student achievement in education, in diagnosing disease symptoms in medicine, determining the associated diseases [9], in smart houses with sensors or in other applications requiring the user's environment to be monitored (such as gas leaks, fire, explosions) [10], and in the analysis of the conditions under which the accidents took place, as well as in the follow-up of daytime activities of dementia patients. The ability to monitor the activities of daily living in the intelligent environment is seen as an important approach to monitoring functional decline among dementia patients [11].

As data mining techniques are applied on millions of data in real life applications, the algorithms used during the association-relationship question must be fast. Market basket analyzes provide data on customer behavior based solely on purchases. This data does not include customer characteristics but provides data on customer distribution at different perspective.

It is attempted to identify common factors in men and women with cardiac disease using the Apriori Algorithm [12]. Palaniappan and Awang [13] studied that the preliminary detection of cardiac risk by using association and classification methods together.

Wen et al. [14] have worked on the improvement of CRM by developing new products, effective promotion studies and sales strategies using the Apriori Algorithm. To analyze the luxury brand consumption habits in Taiwan, the association rule approach and clustering analysis were used to investigate the relationship between luxury consumer products and customers purchasing these products. Ay and Çil [15] have studied shelf management using the Apriori Algorithm. It develops a relational database and uses Apriori Algorithm and multidimensional scaling techniques as methods for the store layout issue.

Budak et al. [16] have worked on the improvement of a web site by identifying visitor needs of a website with in site search by using Apriori Algorithm. Sabnani et al. [17] have devoloped a recommendation system based on association methods to help students to take decisions on their academic itineraries. More specifically, it provides support for the student to better choose how many and which courses to enroll on, having as basis the experience of previous students with similar academic achievements. Tsai and Sheng's experimental study results show the proposed method can reassign items to suitable shelves and dramatically increase cross-selling opportunities for major and minor items [18]. Chen and Lin studied on shelf management, in their study the multi-level association rule mining is applied to explore the relationships between products as well as between product categories [19].

In addition, it is possible to find many examples of fieldwork in the credit risk assessment stages of banks, especially in the detection of frauds on web transactions, in the detection of disease symptoms in the medical field, and in the sequential purchasing of customers (such as insuring your car after buying a car).

3. METHOD

The CRISP-DM (CRoss-Industry Standard Process for Data Mining – CRISP-DM) model process steps were followed in the developed application related to the CRM that has been improved with association rules [20]. In the view of such steps addressed in the data mining which it is used in our application related to CRM; Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment are all explained.

Business Understanding

Businesses track customer purchasing movements during the process of defining customer habits. In this point, they utilize from association methods. Based upon the purchasing movements, answers are sought for the questions such as which products are sold together with which products, and what similarities exist between customer groups that buy these products.

In sales / marketing campaigns organized by the firm, it is very important to determine which group of customers is going to be directed and also which group of products is coming to the forefront in the process of developing a solution proposal on how to combine the best-selling product with the least-sold product. From this point forth, the basic point to be addressed in the application will be to discover which dual, triple, quadruple combinations from the products are purchased together in the company's product range. In this way, these associations can be taken into consideration in marketing campaigns, new sale targets can be created on a product basis by bringing together the products that are sold more and the products that are less sold, so that these product combinations may be included in the catalogs to be submitted to the customers. Likewise, the customers whose purchasing behaviors were examined can be divided into groups. Thus, different sales policies can be offered to the customers in different groups. By identifying the risky customers and also the customers who meet most of the revenue from payment statements, it is possible to make progress in the light of this information on customer relations and sales policies to be developed.

Data Understanding

Data set used in the analyses was provided by the X-firm operating in the electricity sector. The data covers the years 2014-2015. The customers are distinguished by the "Current Code" area. Analyses are made on a yearly basis and are considered comparatively. The first part of the data set belongs to the period January 2014-

December 2014, with a total of 19 attributes and 177393 records. The second part belongs to the year 2015 and consists of 23 attributes and 183401 records. The attributes are shown in Table 1.

Table 1. Attributes of dataset used in analyses

Attribute Name	**Explanation**	**Data Type**
Invoice_No	Invoice number	Numeric
Date	Date of sale	Date
Customer_code	Customer code	Numeric
Product_code	Product code	Categorical
Product_name	Product name	Categorical
Brand	Product brand	Categorical
Brand2	Revised product brand	Categorical
Product_subgroup	Product_subgroup	Categorical
Cable Type	4 different types of cable	Categorical
Valor_day	The day on which interest calculation starts	Numeric
Expiry_date	Expiry date	Date
Discount_rate_1	Discount rate	Numeric
Total_discount	Total_discount	Numeric
Warehouse	Product warehouse	Categorical
Unit	Product unit	Categorical
Unit_quantitiy	Product quantitiy	Numeric
Unit_price	Product unit price	Numeric
Unit_amount	Product unit amount	Numeric
Premium	Premium	Categorical
Typepcode	Salesman	Categorical
Groupcode	Satış region	Categorical
Net_unit_price	Product net unit price	Numeric
Net_amount	Product net amount	Numeric

Data Preparation

Missing Values

The data set was checked for missing values and 39 records were found in which the group code and type code areas were left blank. Since the Group-code and Typecode areas are categorical area types, it is possible to fill them with the most repeated area. However, by communicating with the company, it has been determined that these recordings belong to the “Illumination Region”, and then the current 39 records have been included in the “Illumination Region”. The Typecode of the “Illumination Region” has been determined as "Servet", and the Typecode areas have been updated in this way.

Three records with empty unit area have been detected. This categorical area has been completed with reference to the Cable Type area which will help us in determining the product type. It has been determined that unit quality is entered as ‘piece’ in the products of which Cable Type is general, and 3 records with empty unit area are completed as being ‘piece’, considering as the most frequent feature.

After the completion of the missing values in the dataset, no changes were found in the analyses. This is thought to be due to the fact that “Illumination Region” represents a small group.

Outliers

A record of which only Net Amount attribute is full but of all other qualities left blank is removed from the dataset, so the number of observations decreased from 183403 to 183402. It has been studied with this observation number in the analyses.

Dublicated observations

No duplicated observation was found in the data set. For this reason, there was no deductions from this dataset at this point.

Modeling

In the study, Apriori Algorithm for associations was applied in order to identify product associations and to carry out strategy and sales campaigns in this direction. By interpreting the rules derived from this model, product associations and campaigns that can be developed are presented. The following is the working principle of the Apriori Algorithm.

Apriori property: All nonempty subsets of a frequent itemset must also be frequent. The Apriori property is based on the following observation. By definition, if an

itemset *I* does not satisfy the minimum support threshold, *min sup*, then *I* is not frequent;

"How is the Apriori property used in the algorithm?" To understand how is the Apriori property is used in the algortihm, we should focus on how L_{k-1} is used to find L_k for $k \geq 2$. A two-step process is followed, consisting of join and prune actions [5].

Steps to Perform Apriori Algorithm

1. The join step: To find L_k, a set of candidate *k*-itemsets is generated by joining $L_{k-1}$with itself. This set of candidates is denoted C_k Let I_1 and I_2 be itemsets in L_{k-1}. The notation I_i [*j*] refers to the *j*th item in I_i (e.g., I_1[*k*-2] refers to the second to the last item in I_1). By convention, Apriori assumes that items within a transaction or itemset are sorted in lexicographic order. The join, L_{k-1} on L_{k-1}, is performed, where members of L_{k-1} are joinable if their first (*k*-2) items are in common. That is, members I_1 and I_2 of L_{k-1} are joined if (I_1 [1] = I_2 [1]) ^ (I_1 [2] = I_2[2]) ^…^ (I_1 [*k*-2] = I_2 [*k*-2]) ^ (I_1 [*k*-1] < I_2 [*k*-1]). The condition I_1 [*k*-1] < I_2 [*k*-1] simply ensures that no duplicates are generated. The resulting itemset formed by joining I_1 and I_2 is I_1 [1], I_1 [2], …, I_1 [*k*-2], I_1 [*k*-1], I_2 [*k*-1].

2. The prune step: C_k is a superset of L_k, that is, its members may or may not be frequent, but all of the frequent *k*-itemsets are included in C_k A scan of the database to determine the count of each candidate in C_k would result in the determination of L_k, (i.e., all candidates having a count no less than the minimum support count are frequent by definition, and therefore belong to L_k). C_k, however, can be huge, and so this could involve heavy computation. To reduce the size of C_k, the Apriori property is used as follows. Any (*k*-1)-itemset that is not frequent cannot be a subset of a frequent *k*-itemset. Hence, if any (*k*-1)-subset of a candidate *k*-itemset is not in L_{k-1}, then the candidate cannot be frequent either and so can be removed from C_k [5].

Association rules let I = {I_1, I_2,…,I_m} be a set of items. Let *D*, the task-relevant data, be a set of database transactions where each transaction *T* is a set of items such that $T \subseteq I$. Each transaction is associated with an identifier, called TID. Let *A* be a set of items. A transaction *T* is said to contain *A* if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subset B$, $B \subset I$, and $A \cap B = \emptyset$. The rule $A \Rightarrow B$ holds in the transaction set *D* with support *s*, where *s* is the percentage of transactions in *D* that contain $A \cup B$. The rule $A \Rightarrow B$ has confidence *c* in the transaction set *D*, where *c* is the percentage of transactions in *D* containing

A that also contain *B*. This is taken to be the conditional probability, *P*(*B*|*A*). That is,

$$\text{support}(A \Rightarrow B) = P(A \cup B) \quad (1)$$

$$\text{confidence}(A \Rightarrow B) = P(B|A) = \frac{\text{support}(A \Rightarrow B)}{\text{Support}(A)} \quad (2)$$

Where;

Support: The support of a rule indicates how frequently the items in the rule occur together.

Confidence: The confidence of a rule indicates the probability of both the antecedent and the consequent appearing in the same transaction. Rules that satisfy both a minimum support threshold (*minsup*) and a minimum confidence threshold (*minconf*) are called strong. By convention, we write support and confidence values so as to occur between 0% and 100%, rather than 0 to 1.0.

Beside support and confidence, there is another interestingness measure that the rule $A \Rightarrow B$ has. It is called as lift. Lift is a simple correlation measure that is given as follows. The occurrence of itemset A is independent of the occurrence of itemset B if P(A U B) = P(A)P(B); otherwise, itemsets A and B are dependent and correlated as events. This definition can easily be extended to more than two itemsets. The lift between the occurrence of A and B can be measured by computing

$$\text{Lift}(A \Rightarrow B) = \frac{P(A \cup B)}{P(A)P(B)} \quad (3)$$

Where;

Lift: indicates the strength of a rule over the random co-occurrence of the antecedent and the consequent.

If the resulting value of equation for lift is less than 1, then the occurrence of A is negatively correlated with the occurrence of B. If the resulting value is greater than 1, then A and B are positively correlated, meaning that the occurrence of one implies the occurrence of the other. If the resulting value is equal to 1, then A and B are independent and there is no correlation between them [5].

The association method preferred in this study; Apriori is a seminal algorithm proposed by R. Agrawal and R. Srikant in 1994 for mining frequent itemsets for Boolean

association rules. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent itemset properties. Apriori employs an iterative approach known as a level-wise search, where k-itemsets are used to explore (k+1)-itemsets. First, the set of frequent 1-itemsets are found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted L1. Next, L1 is used to find L_2, the set of frequent 2-itemsets, which is used to find L_3, and so on, until no more frequent k-itemsets can be found. The finding of each L_k requires one full scan of the database.

minsupp and minconf values are given by users or experts. Then A⇒B is a valid rule

If

$$\text{support}\,(A \Rightarrow B) \geq \text{minsupp}, \quad (4)$$

$$\text{confidence}\,(A \Rightarrow B) = \frac{\text{supp}(A \cup B)}{\text{supp}(A)} \geq \text{minconf} \quad (5)$$

Mining association rules can be broken down into the following two subproblems:

(4) Generating all itemsets that have support greater than, or equal to, user specified minimum support. That is, generating all frequent itemsets.

(5) Generating all rules that have minimum confidence in the following simple way: For every frequent itemset X, and any B⊂X, let A = X − B. If the confidence of a rule A ⇒ B is greater than, or equal to, the minimum confidence, then it can be extracted as a valid rule [21].

To sum up; the support value is calculated as the ratio of the number of records in which A and B are together to the total number of records in the database [22]. This criterion indicates the frequency of the relation between the items in the data. The confidence value of A ⇒B rule represents the power and reliability of the inference (rule) [23]. The rules obtained from the analysis are interpreted according to these interestingness measures at findings and discussion section.

Apriori Algorithm has some limitations as follows: (1) for generating candidate sets, Apriori Algorithm requires number of scans over the database. Because of the multiple scanning over database it takes lots of time to execute and increases I/O load. (2) Also, while scanning databases it generates the number of candidate sets in database. In order to overcome the drawbacks of Apriori Algorithm, there are various types of improvement in this algorithm like matrix, weighted Apriori, hash structure; interest itemsets, transaction compression etc. are possible.

In this study, analyses were performed with the R programming language. R is a programming language and environment for statistical computing and graphics [24]. In this study, RStudio, the visualized and integrated development environment of R, was used. The arules [25] library on R was utilized to obtain the association rules that can be attained with Aproiri. The arulesViz [26] library was used to visualize the rules. The application of the Apriori Algorithm on Shiny has been developed.

Shiny is R's web application framework [27]. In this way, it has been enabled that the developed model can be easily applied on the web by the end user [24]. In order to interpret the model obtained from the association rule, the support and confidence values reached as a result of the rule are utilized. The support and confidence values of the rule are two measures that express the interestingness of the rule. These values express respectively the efficacy (usefulness) and the certainty (correctness) of the discovered rules [28]. The steps described in modeling are covered in the findings and discussion section

4. FINDINGS AND DISCUSSION

In the Shiny application, the association analysis performed with the Apriori Algorithm has been moved to the web environment. Thus, regardless of whether the program is installed on their computers, the application has been not made accessible and practicable from everywhere not only to analysts but also to each employee (field authority, marketing department staff, department managers, product/region representatives etc.) within the institution who will be able to use the analysis results.

To reach the application online, you can follow the link (https://gkarahanadali.shinyapps.io/shinyapriori3/), the server should be on running mode to access the program [29].

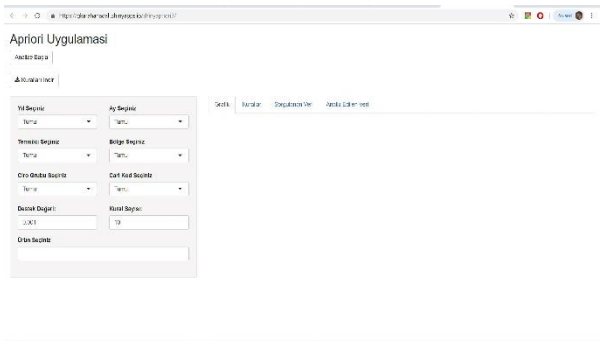

Figure 1. Shiny Application Interface

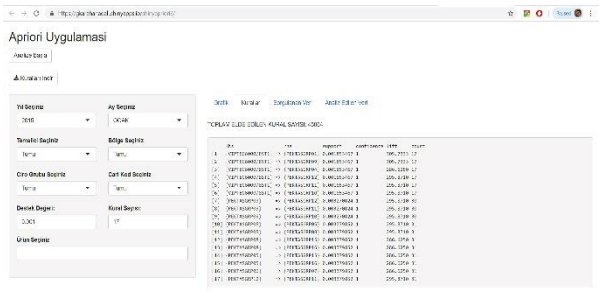

Figure 2. Displaying the Rules by Shiny Application

Thanks to the Shiny application, this process is made possible by leaving it to the user's preferences. It is aimed that the user can reach the results quickly from the right way by passing the analysis phase. The data set that the user wants to analyze is left to his / her own choice by designing a simple, modest and user-friendly interface allowing the user to filter on the data set. Scatter plots of created rules can be displayed as well. Figure 1 shows the welcome page of the program. By using the left panel: The user can filter the by choosing the year, month, region, endorsement group of customer and customer_id. In addition, the user can assign the support value and the shown rule number and also can select a specific product. The user can start analyse by submitting the Start Analyse button. When the analyse is completed, the user can download the rules on his/her desktop. So, in the future steps, rules can be compared to obtain best result.

Segmentation is defined as the separation of the customer base into different groups in order to develop differentiated marketing strategies according to their own characteristics [1]. In this study, with the help of pareto analysis, customers are divided into groups A, B and C according to the turnover they have made during the year. In this way, it is possible to analyze the data through these turnover groups of which sales were determined. By deducting the purchasing habits of these turnover groups, it can be thought that the customer who is in the same group and buys X, Y, Z products is a reference for another customer who is in the same group and buys only X, Y products, and is potential buyer of Z products.

In the study, the pruned rules mentioned in Modeling section are interpreted. The data set covering both years is analyzed with the support = 0.001 and confidence = 0.5 threshold values, it draws attention that the HES BRAND cable has a visible sales association in its products as scattered on all the months except December and January. When December and January are examined, due to the fact that the support, confidence and leverage values of the PEKTASGRP09 }=>{PEKTASGRP10} product pair were observed as support value= 0.001, confidence = 1 and lift = 887.8000 respectively, it has been deemed appropriate to present these two products as a package for customers.

The rules are shown in Rules tab to the user (see Figure 2). User can also see the queried data (consist of all attributes) and analyzed data (only the attributes that is necessary for the analyses) which are different from each other. The code prunes the data at the back while running. The rule mentioned above is interpreted as follows:

All of those who buy PEKTASGRP09 (confidence) product also buy PEKTASGRP10 product. The purchase of those who buy these two products together makes up 0.1% of all purchases (support). Because the lift value is 887.800> 1, it can be said that this rule is reliable, and its performance is high.

In the product preferences, when PEKTAS product group is examined by considering net sales amounts at the end of year as December and the beginning of year as January, the decrease in other major brands' sales at the end of the year compared to previous year can be interpreted as the company’s attempt to finish stocks on hand. With the withdrawal of major players on the market, other brands have been on the forefront these months.

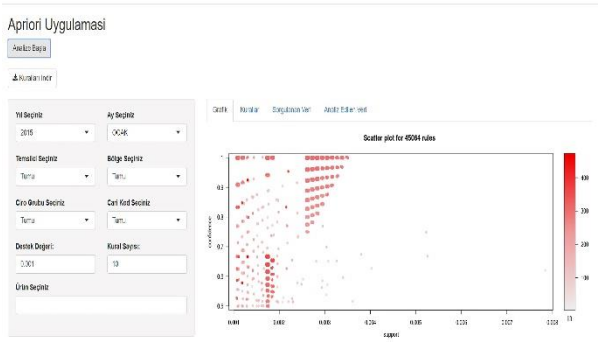

Figure 3. A scatter plot for 2015; January

In Figure 3, a scatter plot is shown according to January 2015. The point on the plot represents the rules. The x axis on the plot represents the support value, the y axis represents the confidence value. The color's tone emphasizes the lift value. In the study, when the customer segments separated as A (highest) -B (medium) -C (lowest) are analyzed on the entire data set regardless of the year in order to measure the association in itself, the following rules have been reached: When ranked according to the confidence values in the 100 rules obtained for the turnover group A, the highest 5 rules are as follows. Those who buy HESHFNYA02-black / green and HESHFNYA03-BLUE products also purchase HESHFNYA02-BLUE products with a probability of 92% (confidence).

The purchase of this triple product package account for 0.1% of all purchases. If the second rule is taken into consideration, those who buy KONKV101, KONKV101,5, KONKV105 products of Kondaş Brand, also buy KONKV102,5 product with 92% probability. The turnover group A is the upper segment of the customers. It is seen that the top customer group concentrates on Kondaş and Hes Brand. When the data set is examined on a yearly basis, the outstanding difference between 2014 and 2015 was the association in Legrand Brand BULEG638038 product seen in 2015.

The dominance of HES Cable products was broken with this product in 2015. It is observed that those who buy BULEG638038 and BULEG674420 also buy BULEG77213 product with 89% probability. 18 regions where the firm has worked are included separately in the analysis. When Karaköy, Trakya2 and Trakya1 Regions, which have the highest turnover, are taken into consideration, the results are as follows:

In the Karaköy Region, which has the highest total turnover in 2014 and 2015, the weight of Hes Cable is observed. The strong association between HESHFNYA02 and HESHFNYA03 products is remarkable. It is seen that those who buy HESHFNYA02-KAH, HESHFNYA03-KAH, HESHFNYA03-MAV products definitely (100%) purchase HESHFNYA02-MAV product. The purchase of those purchasing this quaternary product package accounts for 0.1% of all purchases. This rule with a leverage value of 185.0291 can be said to be highly reliable.

These two products can be transformed into a color palette and then presented to the customers so that it is anticipated that other products can be offered in addition to this package to enable customers to purchase these additional offered products. When the Trakya2 Region, which is second in the turnover rank, is examined, it is observed that the association between Kondaş Brand product groups is remarkable. When the obtained rules are ranked according to the confidence value, if necessary to put it in the first place, it is seen that the buyers of KONKV101,5 KONKV105, KONKV107,5 products also purchase KONKV 102,5 product. The purchase of those purchasing this quaternary product package accounts for 0.1% of all purchases [30].

When the Trakya1 region, which is in the third place in turnover order, is examined, it was deemed appropriate to create a color palette between HESHFNYA02 and HESHFNYA03 products in HES Brand cables. The selections can be reproduced in this way. One of the aims desired to obtain from the Shiny application is to enable the user to perform a deep analysis of the data at hand by filtering on the data set selected by him/herself.

Through the developed model, product analyses can be performed by the company authorities or the related end user without any software knowledge. Thanks to the interface presented to the user, there is a chance to intervene in the data set. The user repeatedly shapes the data in the direction of the values that he / she has selected. Thus, the application becomes dynamic. The user can select the desired year from the data set that scattered into two years or analyze it as in the form of covering all the years.

By Shiny application, it is possible to examine 18 different regions in the dataset individually. The association in the product sales of defined turnover groups can be examined in line with the elections made. By focusing much more on a specific area, monthly sales data in the area of a selected sales representative can be examined. When it is desired to conduct a study on a specific product, it has been made possible to examine all the rules concerning a product in the rule. The number of screen views of the generated rules is presented to the user's preference. In order to be able to view and examine the created rules in more detail, the

user has the option of downloading the analysis results to his/her computer. In the Apriori Algorithm, the support threshold value defined once by the user has been made dynamic. In the case where the rules created are too large or too small to be analyzed, the user has been instructed that he/she needs to change the support threshold.

Especially the association of most common HES Brand cable sales among the obtained results is remarkable. The cross sales realized between HESNYA-03 and HESNYA-02 products of this brand and the sales of different colors together bring the product into the forefront. From this result, it is thought that by creating various color packages related to the product, customers can increase their purchases of less preferred products by being offered new products besides these and similar packages. Consequently, by using the association analysis from data mining context, it has been made also possible to apply the possible product packages on a selected customer or customer groups.

5. CONCLUSION

In this study we have presented an association rule mining application developed by Shiny web interface. Running of the model, which was created through the developed web interface, at the back face allows the application to have a dynamic feature independent of time and space. The developed model can be easily applied by the end user on the web without program dependency. In this way product analyzes can be performed by the company authorities or the related end user without any software knowledge. Under favour of the interface presented to the user, there is a chance to intervene in the data set. The user shapes the data repeatedly in the direction of the values that he / she has chosen.

The user is given the opportunity to query the analyzed data set and make basic arrangements related to the algorithm. This allows the application to be dynamic, independent of time and space.

The cross sales realized between HESNYA-03 and HESNYA-02 products of this brand and the sales of different colors together bring the product into the forefront. From this and suchlike results, it is thought that by creating various color packages related to the product, customers can increase purchases of less preferred products by being offered new products besides these and similar packages.

The data set used in the study belongs to a company operating in the wholesale sector. This is an important limitation of the analysis. In further studies, it is recommended that the Apriori Algorithm should be used in the retail sector and with a larger data set for more effective results. Analyzes were applied on an annual and monthly basis in the Shiny application. For the more detailed results, the daily basis is recommended.

In this study; it is thought that this model, which is designed to collect sales and increase sales, can make faster decisions for customers and sales by running R model on the web. By the developed model, the dynamic applicability of possible product packages on a customer or customer groups in line with the user preferences has been enabled. A significant contribution has been made to the literature by developing a dynamic model that is considered to enable the user to conduct customer and product analysis so that can provide competitive advantage. It is aimed that this study can be applied in different fields of the data mining by adding different models to be applied in the sector.

REFERENCES

[1] M. Karabatak, M. C. İnce, "Apriori Algoritması ile Öğrenci Başarısı Analizi", **Eleco`2004 Elektrik - Elektronik - Bilgisayar Mühendisliği Sempozyumu ve Fuarı**, http://www.emo.org.tr/ekler/24f4c5eef7ec01c_ek.pdfursa, 2004.

[2] D. Birant, A. Kut, M. Ventura, H. Altınok, B. Altınok, E. Altınok, M. Ihlamur, "İş Zekası Çözümleri için Çok Boyutlu Birliktelik Kuralları Analizi", 215–222, 2010.

[3] K. Tsiptsis, A. Chorianopoulos, **Data Mining Techniques in CRM**, 2009.

[4] T. C. Yang, H. Lai, "Comparison of product bundling strategies on different online shopping behaviors", *Electronic Commerce Research and Applications*, 5(4), 295–304. doi: 10.1016/j.elerap.2006.04.006, 2006

[5] J. Han, M. Kamber, **Data mining: Concepts and techniques**. San Francisco: Morgan Kaufmann Publishers, 2006.

[6] C. F. Özçakır, **Müşteri İşlemlerindeki Birlikteliklerin Belirlenmesinde Veri Madenciliği Uygulaması**, Yüksek Lisans Tezi, Marmamara Üniversitesi, Fen Bilimleri Enstitüsü, 2006.

[7] Internet: S. Erpolat, Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması, https://earsiv.anadolu.edu.tr/xmlui/handle/11421/163, 2012.

[8] P. Kumar, "Knowledge Discovery in Databases (KDD) with Images : A Novel Approach toward Image Mining and Processing", *International Journal of Computer Applications*, 27(6), 10–13, 2011.

[9] D. Delen, G. Walker, A. Kadam, "Predicting breast cancer survivability: A comparison of three data mining methods". *Artificial Intelligence in Medicine*, 34(2), 113–127, https://doi.org/10.1016/j.artmed.2004.07.002, 2005.

[10] M. M. Rashid, I. Gondal, I., J. Kamruzzaman, "Mining

associated sensor patterns for data stream of wireless sensor networks", **Proceedings of the 8th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks - PM2HW2N '13**, New York, NewYork, USA, ACM Press, 91–98, https://doi.org/10.1145/2512840.2512853, 2013.

[10] R. Agrawal, J. C. Shafer, **Parallel mining of association rules: Design, Implementation and Experience**, IBM Research Report RJ 10004, 1996.

[11] S. Nasreen, M. A. Azam, K. Shehzad, U. Naeem, M. A. Ghazanfar, "Frequent pattern mining algorithms for finding associated frequent patterns for data streams: A survey", *Procedia Computer Science*, 37, 109–116, doi: https://doi.org/10.1016/j.procs.2014.08.019, 2014.

[12] J. Nahar, T. Imam, K. S. Tickle, Y. P. P. Chen, H. H. Yang, S. Fong, Y. Zhao, "Association rule mining to detect factors which contribute to heart disease in males and females", *Expert Systems with Applications*, 102(3), 335–351, doi: https://doi.org/10.1016/j.jss.2014.07.010, 2013.

[13] S. Palaniappan, R. Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", *International Journal of Computer Science and Network Security*, 8, 343-350, 2008.

[14] C. H. Wen, S. H. Liao, W. L. Chang, P. Y. Hsu, "Mining shopping behavior in the Taiwan luxury products market", *Expert Systems with Applications*, Elsevier Ltd, 39(12), 11257–11268, doi: 10.1016/j.eswa.2012.03.072, 2012.

[15] D. Ay, İ. Çil, "Migros türk a.ş.de birliktelik kurallarinin yerleşim düzeni planlamada kullanilmasi", 14–29, 2010.

[16] V. Ö. Budak, E. Kartal, S. Gülseçen, "Site-içi Aramalar ve Apriori Algoritması Kullanılarak Web Sitesi Ziyaretçilerinin İhtiyaç Tespitine Yönelik Bir Örnek Olay İncelemesi", *Bilişim Teknolojileri Dergisi*, 11(2), 211-222, 2018.

[17] H. Sabnani, M. More, P. Kudale, S. Janrao, "Prediction of Student Enrolment Using Data Mining Techniques", *International Research Journal of Engineering and Technology (IRJET)*, 5(4), 1830-1833, 2018.

[18] C. Y. Tsai, H. Sheng, "A data mining approach to optimise shelf space allocation in consideration of customer purchase and moving behaviours", *International Journal of Production Research*, 53(3), 850-866, doi: https://doi.org/10.1080/00207543.2014.937011, 2014.

[19] M. C. Chen, C. P. Lin, "A data mining approach to product assortment and shelf space allocation", *Expert Systems with Applications*, 32(4), 976–986, doi: 10.1016/j.eswa.2006.02.001, 2007.

[20] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining", *Journal of Data Warehousing*, 5, 2000.

[21] C. Zhang, S. Zhang, **Association Rule Mining Models and Algorithms**, Berlin: Springer, ISBN 3-540-43533-6, 2002.

[22] P. N. Tan, M. Steinbach, V. Kumar, "Association Analysis: Basic Concepts and Algorithms", *Introduction to Data Mining*, 327–414, doi: https://doi.org/10.1111/j.1600-0765.2011.01426.x, 2005.

[23] R. Kaur, "Apriori algorithm for Mining Frequent Patterns using Parallel Computing : Survey", *International Journal*, 6(5), 822-824, 2016.

[24] Internet: R: The R Project for Statistical Computing. (n.d.), https://www.r-project.org/, 05.01.2016.

[25] Internet: M. Hahsler, C. Buchta, B. Gruen, K. Hornik, I. Johnson, C. Borgelt, arules: Mining Association Rules and Frequent Itemsets, https://cran.r-project.org/web/packages=arules, 2019.

[26] Internet: M. Hahsler, G. Tyler, S. Chelluboina, arulesViz: Visualizing Association Rules and Frequent Itemsets, https://cran.r-project.org/web/packages/packages=arulesViz, 2018.

[27] Internet: RStudio – Open source and enterprise-ready professional software for R. (n.d.), https://www.rstudio.com/, 05.01.2016.

[28] M. E. Balaban, E. Kartal, **Veri Madenciliği ve Makine Öğrenmesi**, Çağlayan Kitabevi, ISBN: 978-975-436-089-9, 2015.

[29] Internet: Shiny. (n.d.), http://shiny.rstudio.com/, 10.01.2017.

[30] G. Karahan Adalı, **Veri Madenciliğinde Birliktelik Yöntemleri Ve Müşteri Ilişkileri Yönetimine Ilişkin Bir Uygulama**, Doktora Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, 2017.