



Operations Research Society in
Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing
Available online: <http://scjournal.ius.edu.ba>



IUS Soft Computing
Research Group

Determination of Association Rules with Market Basket Analysis: An Application in the Retail Sector

Ayşe Nur Sağın^a, Berk Ayvaz^b
Graduation School of Natural and Applied Sciences,
Istanbul Commerce University,
Küçükyalı E5 Kavşağı İnönü Cad. No: 4,
Küçükyalı34840, İstanbul
^aanur.sagin@ticaret.edu.tr,
^bbayvaz@ticaret.edu.tr

Article Info

Article history:

Article received on 17 January 2018
Received in revised form
10 February 2018

Keywords:

Apriori algorithm, Association rules,
Data mining, FP-Growth algorithm,
Market basket analysis

Abstract

Market basket analysis is the process of extracting purchasing trends from records in company databases, taking into account the products that customers buy in a single transaction. In this study, a market basket analysis was conducted on a five-and-a-half year data of a large hardware company operating in the retail sector, and related product categories were identified. In determining the association rules, both the Apriori and FP-Growth algorithms were run separately and their usefulness in such a set of data was compared. In addition, the data set was divided into Data Set-1 and Data Set-2 so that the consistency of the rules was discussed by comparing the correctness of rules extracted from the first data set with rules derived from the second data set containing consecutive timed data.

1. INTRODUCTION

All kinds of transactions that we perform in daily life are recorded in digital environments thanks to advanced data base systems. The information of the products we buy together in a grocery store and those that we return, banking transactions, phone calls or any transactions we make during online shopping are recorded simultaneously [1]. Such a development in information technology has also brought some problems for companies. In businesses with large numbers of customers and transaction activity, huge amounts of data are being piled up each day due to the production of large amounts of data. Today's database technology has developed enough to keep these data stacks healthy, but it is important not to keep those data, but to evaluate the data to provide added value for the company. Businesses need to develop effective and low-cost

marketing strategies so that they can survive competitive conditions in today's consumer-focused markets [2]. In order for strategic decisions to be made, reliable and, as much as possible, evidence-based knowledge is needed. With advancing technology, data mining has become one of the best answers to this need. Data mining is the technique of filtering valuable information from large databases[2]. Data mining makes it possible to transform confidential information in large data stacks into valuable information for the company.

In this study, the determination of association rules, which is one of the methods of data mining, was realized by market basket analysis. The analysis was carried out on sales data from a branch of a major hardware company. The purpose of the analysis is to identify the product groups that tend to be sold together within a certain period determined in the

business. To test the validity of the rules to be derived at the successive periods was divided into two parts. The data set was passed through various stages and transformed into a format that WEKA, a data mining software, can understand. Using the Apriori and FP-Growth algorithms in the WEKA program, the rules of association in the data set were derived. The performance of the algorithms on a data set of this size was evaluated.

The subsequent parts of the article are structured as follow: In the second part, a literature review involving the purpose of the use of association rules in various sectors is included. In the third part, data mining, which is the main topic of the study, and the operation of the algorithms used in the study are given. In the fourth part, the implementation steps are explained. The fifth chapter, the last chapter, contains the results of the study.

2. LITERATURE REVIEW

Poel et al. determined the associated product pairs by using the market basket analysis on a yearly data from a "Do it yourself" (DIY) retailer. The effects of various promotional activities on these complementary product pairs were analyzed and the necessary recommendations were made in the direction of the results[3]. Chen et al. assumed that, in enterprises with chain stores, major purchasing patterns could not be extracted because of the assumption that products are always on store shelves at all times. In order to overcome this problem, they also added warehouse and time information to the rules developed differently in similar studies. Thus, they aimed to use the association rules in the development of marketing strategy, product supply, inventory and distribution strategies for the entire store chain[4]. Yun et al. believed that the traditional association rules of mining could not produce healthy results in multi-store environments. An algorithm similar to the Apriori algorithm was developed to examine the association rules in these stores. The rules generated by the developed algorithm also provided information about shop location and time period. It was seen that the proposed method was better than the traditional Apriori algorithm in terms of simulation work done in shops with varying sizes and continuously changing product mixes[5]. Kuo et al. worked with the Taiwan National Health Insurance Agency to find out the rules of association among diseases on the twelve-month health insurance data. In their study, they utilized the ant colony system, clustering and association rules[6]. Erdem and Özdağoğlu conducted an association rule analysis on the emergency services of an education and research hospital in the Aegean Region. According to the rules of association found using the Apriori algorithm, it was observed that there were associations regarding gender, stay time, emergency service arrival time, and some disease diagnoses. This study is expected to guide the reconstruction of emergency departments by determining the causes of emergency service applications of the patients in that region and the patient

end of the study, the data set containing the data in profiles[7]. Ergün used association rules and clustering analysis to identify related product categories in a retailer's store. Store products were grouped as non-food, fast-food and ready-to-eat food and their inter-subcategory relations were analyzed using customer receipt data. Thus, the associated categories were identified[8]. Timor and Şimşek determined factors affecting customers' buying behavior with decision trees in one of Turkey's largest chain of markets operating in the retail sector, by analyzing association rules on customer shopping records for a period of four months in 2004[9].

Li and Li proposed a new system of alarm correlation analysis based on association mining to analyze the relationships between failures in a telecommunication company[10]. In his work on three different internet shopping markets in Taiwan, Chiang used the supervised Apriori algorithm to learn valuable customers and developed the RFMDR model[11]. Nahar et al. used association rules to identify factors that cause bladder, breast, prostate, cervical, lung, and skin cancers. Among the three different algorithms used for this purpose, they found that Apriori was the best contributor[12]. In his study, Durdu, conducted the association analysis on the products in the food category of Pehlivanoglu Marketing Food Trading Inc. and determined the associated categories by using the Apriori algorithm[13]. In his study, Erpolat presented information about the Apriori and FP-Growth algorithms, which are widely used in the analysis of association rules, applied these algorithms on the customer shopping data of an authorized automotive service and compared the results. As a result of the study, he found that FP-Growth algorithm is more suitable than Apriori algorithm in analyzing a data set of this size[14]. Cil offered a new proposal for store placement problem with association rules and multidimensional scaling analysis. In his research, he took customer receipt and product barcode information as data for the MigrosTürk supermarket, which has an important place in the retail sector, and analyzed it using the Apriori algorithm, according to the method of association rules. He then suggested a new store layout plan based on the resulting rules[15]. Hanguang and Yu worked on a Web based attack detection system with association rules mining and Apriori algorithm [16]. Güngör et al. n.d. conducted association rule analyses using Apriori algorithm on the questionnaire data with the aim of determining what criteria students use when selecting technical elective courses in Computer Engineering Department of BilecikŞeyhEdebalı University[17]. Mansur and Kuncoro used market basket analysis and artificial neural networks to examine customer buying behaviors in a company selling furniture accessories called Karomah Brass. With this study, they obtained useful results on inventory planning for the company[18]. Paranjape-Voditel and Deshpande, implemented the association rules mining on the stock market sector. They developed a portfolio recommendation system with market

basket analysis[19]. Dhanabhakya and Punithavalli recommended the FRG-AARM approach for an effective market basket analysis. In this proposed approach, unlike the traditional approach, the best features of the Apriori and Apriori-Tid algorithms were combined with the Apriori-Hybrid algorithm. The model suggested by the simulation study was assessed with the traditional models in terms of prediction accuracy and average run-through performance. It was seen that the proposed approach yielded results in a very short period of time[20]. Nahar et al. identified significant risk factors for heart attacks in men and women with the rules of association[21]. Ceylan, in his work in a pharmacy, used patient prescription data to analyze associations between drugs. As a result of the analysis, shelf regulation was carried out with the rules determined. With the new shelf scheme, pharmacist's drug search time decreased and thus customer satisfaction increased. At the same time, this system aims to prevent the possibility of faulty drug delivery in pharmacies[22]. Doğan et al. emphasized the idea that increasing the loyalty and satisfaction of customers in an insurance company by determining the purchasing habits of the customers could increase the profit. As a result of the analysis of the association with the Apriori algorithm, five prominent rules were identified[23]. In their study performed in the stock market database, Kamley et al. aimed to determine the investment tools that would bring maximum profits to the stock market and investors and determine the appropriate time[24].

Kaur and Kang, suggested an algorithm that would perform association rule mining on dynamic data. By working with the concept of change modeling, they performed periodic mining[25]. Solnet et al. carried out a market basket analysis on the data of a luxury hotel group in Australia. They aimed to increase the hotel's profit with the selection of the most appropriate supplementary services for the guests[26]. In their research, Yazgan and Kusakci worked on improving the order collection process in a pharmaceutical warehouse. In order to solve this problem, where the products to be stored are small in size, the order stacking strategy was proposed. By determining the rules of association between customer orders, stacks were created with genetic algorithm in order to match the same stack of order receipts with similarity in different pharmaceutical warehouses. With more than one order collection in one round of the order picker, the number of rounds was reduced, which saved time[27]. Özçalıcı used the Apriori algorithm on 211,109 advertisements posted on sahibinden.com, the second hand car sales site, to determine the general features of the vehicles offered for sale. The Apriori algorithm was used for the first time for such a purpose[28]. Yildirim et al. tried to determine the relationship between yarn and fabric characteristics with their association rules mining study performed on the data of a textile company. In determining the association rules, the FP-Growth algorithm was used[30].

3. METHOD

In this study, it is aimed to determine the association rules, one of the data mining methods, with market basket analysis technique. In this part of the study, information on data mining, the main topic of the study, association rules, and algorithms used for this analysis were given.

3.1. Data Mining

All kinds of transactions that we perform in everyday life are recorded in digital environments thanks to advanced database systems, and these records are turning into larger and larger data clusters day by day. As a result of their work in 1992, William et al. estimated that the amount of data on Earth would double in every twenty months [31]. This exponential increase over time causes large data clusters to form. All of the process of converting these data chunks from specific processes to usable information has entered our lives for the first time as 'Discovery of information from the database' with the Knowledge Discovery in Databases (KDD) study conducted in 1989. Data mining is a phase of the VTBK process on a specified set of data, in which various patterns are obtained by the use of methods such as classification, clustering, association rules, and the appropriate algorithms for these methods.

To make a more specific definition for data mining, it is the process of extracting valuable information that is hidden in the huge masses of data unknown beforehand and which we can not predict with our intuition and experience and which will be of great benefit to the business when it is learned. [32].

3.2. Association (Relation) Rules

The association rules used to discover hidden relationships in a large data set are descriptive models of data mining. With this method, a series of operations are applied to the records in the database in bulk, and the rules describing the relations between the records are extracted. By the technique of association rules, it is presented as a summary by evaluating how much the development of a series of events affects the occurrence of a certain event. Rules that are equal to and higher than the success rate set at the beginning of the application are determined by the analyst[33]. With this technique, all possible interesting patterns in the database can be accessed. Thanks to access to all points, the possibility of application even in the growing heaps of data that becomes larger as a result of exponential growth of the records in the database, and the fast calculation logic, it is becoming increasingly widespread in the mining of commercial databases.

The association rules, which are often referred to in the literature as market basket analysis, are used to identify related products that are purchased together in a single trade

or purchased within a certain period of time after a product is purchased. Thus, information can be obtained about customers' cross-purchasing behaviors[34].

Mathematical model of association rules

The mathematical model of the association rule was discovered in 1993 by Agrawal, Imielinski and Swami.

In the model, the set of objects is expressed as $I = \{i_1, i_2, i_3, \dots, i_m\}$, and the set of operations is expressed as D . Every i corresponds to a different object (product). Each transaction in database D is an itemset defined as $T \subset I$. TID is a unique number for each action and m is the number of items. A and B represent object sets. A set of T operations is said to contain A if and only if $A \subset T$, that is, if A is a subset of T . Assuming that A and B are itemsets to satisfy the conditions $A \subset I$, $B \subset I$ and $A \cap B = \emptyset$, then the association rule is expressed as $A \Rightarrow B$. Here A is the premise, B is the result [35].

If the data set in which the basket analysis is performed contains only information on whether or not the product is sold, the correlation between the products is calculated through support and confidence values [36]. The *number of support* is used to denote the number of products containing product groups A and B , and is represented as $number(A \Rightarrow B)$.

The rule of support is used to express the proportion of an association in all exchanges made;

$$Support(A \Rightarrow B) = \frac{number(A, B)}{\text{all shopping}} \quad (1)$$

The rule of confidence establishes the likelihood that customers receiving product group A will also receive product group B . It is calculated as

$$Confidence(A \Rightarrow B) = \frac{number(A, B)}{number(A)} \quad (2)$$

Another criterion in the literature for market basket analysis is 'Lift'. In Turkish sources, this word appears to be met with multiple words such as elevator, interesting, lift, threshold.

$$Lift(A \Rightarrow B) = \frac{Confidence(A \Rightarrow B)}{Support(B)} \quad (3)$$

For example, a lift value greater than 1 between A and B can be interpreted as a positive effect of A on the arrival of, or a positive correlation between them. For all lift values greater than 1[37];

$$\text{Actual Lift} = (\text{Lift value} - 1) \quad (4)$$

,and in this case

$$\text{Percent increase} = (\text{Lift value} - 1) * 100 \quad (5)$$

The interpretation and application of the formula is demonstrated on the part of the results.

Derivation of association rules from databases usually takes place in two stages: [32]

1. Finding frequent itemsets: At this stage, itemsets that provide a predetermined minimum support value are found. This is the step that determines the performance of the association rules. A k -itemset has a $2^k - 1$ number of non-empty sub-sets. All of these non-empty sub-sets can be potential frequent itemsets. In databases where the value of K is large, it is difficult to calculate and keep this process in memory even with computers. Some algorithms have been developed for this. The most common of these, Apriori and FP-Growth algorithms, will be explained in the following parts.

2. Establishment of strong association rules from frequent itemsets: This is the step in which the association rules are established from the frequent itemsets in the first step which establish the minimum trust threshold. It is not a difficult step to formulate the rules of association once the frequent itemsets have been identified.

3.3. Apriori Algorithm

The problem of mining of association rules was initiated by Agrawal in 1993. In 1994, Agrawal and Srikant developed the Apriori algorithm, which is the most well known and used algorithm today. Algorithm is based on predicting the itemsets that are passing frequently from the previous step. The name comes from the word 'prior'. [38]

The Apriori algorithm says that all sub-clusters of an uncommon itemset can not be frequent, or in other words if an itemset does not provide the minimum threshold value, the itemset's parent groups do not provide the minimum threshold value, either.

- k -itemset represents the itemset containing: k number of elements.
- L_k refers to frequent itemsets with: k elements.
- C_k refers to frequent candidate itemsets with: k elements.

The Apriori feature is used to reduce the number of objects that need to be searched to find frequent object sets. This algorithm proceeds in an iterative manner to the operations of finding 2-itemsets using 1-itemsets, and 3-itemsets using 2-itemsets. This can be generalized as follows; $(k-1)$ element

frequent itemsets are used to find k element frequent object candidates. The Apriori algorithm's function is as follows [35];

- First, 1-itemset is found by scanning the database.
- Then, frequent itemsets are found among the 1-itemsets, and are expressed as L_1 .
- After that, 2-itemsets are created using L_1 .
- L_2 is created by determining the frequent occurrences among the 2-itemsets.
- Similarly, the process is repeated so that L_3, L_4 , and so on can be found.

According to the Apriori feature, k-itemsets that do not provide minimum support value are eliminated in each iteration, and only the itemsets providing the threshold support value can be used to generate (k + 1) itemset. This operation continues until the algorithm can not find frequent itemset any longer. Thanks to this process, the number of possible frequent itemsets that need to be scanned when determining the frequent itemsets is greatly reduced.

Table 1. Apriori Algorithm code and Apriori-gene function [39]

```

1)  $L_1 = \{ \text{Large 1-itemsets} \};$ 
2) for (k=2;  $L_{k-1} \neq 0$  ; k++) do begin
3)    $C_k = \text{Apriori-gen}(L_{k-1});$ 
4)   forall transactions - t  $\in D$  do begin
5)      $C_t = \text{subset}(C_k, t);$ 
6)     forall candidates  $c \in C_t$  do
7)       c.count++;
8)     end
9)      $L_k = \{ c \in C_k \mid c.\text{count} \geq \text{minsup} \}$ 
10) end
11) Answer =  $\cup_k L_k$  ;

           Algorithm Basic

insert into  $C_k$ 
select  p.items1, p.items2, p.items3,.....,
p.itemsk-1, q.itemsk-1
from  $L_{k-1}$  p,  $L_{k-1}$  q
where  p.item1=q.item1,.....,    p.itemk-2
      =q.itemk-2 ;
p.itemk-1 < q.itemk-1 ;
forall itemsets  $c \in C_k$  do
forall (k-1) - subsets s of c do
if (s  $\notin L_{k-1}$ ) then
delete c from  $C_k$ 

           Apriori-gene function
    
```

With Apriori-gene function codes given above, C_k s, which are the candidate frequent itemsets, are formed. From the algorithm, it is seen that L_{k-1} frequent itemsets are used to generate C_k from k-itemsets. First, the object L_{k-1} is merged with itself. In the merging process, cross-resemblance of other items is sought in L_{k-1} candidate frequent itemset except for the last element in each line, and candidate frequent itemsets are generated from similarities captured with other items, except for the last element [40]. Generated candidate frequent itemsets are subjected to pruning procedures and reverse the candidate frequent itemset.

In pruning, element deletion is performed from the common element candidate set C_k , if one of the subsets of an item in the candidate object set is not in the common object set L_{k-1} , then this element is deleted from the candidate element object set C_k [41]. This cycle continues until frequent itemset cannot be found.

3.4.FP-Growth (Frequent Pattern Growth) Algorithm

The FP-Growth algorithm is a methodology for finding itemsets without involving the Apriori algorithm's generating and testing steps. When the Apriori algorithm is run on large databases, the database will be scanned hundreds of times, which will lead to huge time loss. With the FP-Growth algorithm, the database is first compressed into a tree structure called the FP tree, where the association information of the itemsets is included. The compacted database is then partitioned into the conditional data structure. Each is associated with a frequent object and these databases are mined separately.

The FP-Growth method transforms the problem of finding large common itemset models repeatedly into the problem of searching the minors and then merging the suffixes. It provides good selectivity by using slightly repetitive objects as a suffix. This method significantly reduces search costs. [42]

The basic execution process of FP-Growth algorithm is as follows [43];

1. First, the database is scanned and items that are equal to and greater than the threshold value are found.
2. The support values of frequent itemsets are listed in order of size (big to small).
3. A tree with only roots is then created.
4. The database is scanned again and for each sample;
 - a) Items taken from the sample are added to the tree, using only the frequent items (items listed in step 2).
 - b) Step a is repeated until all the samples are processed.

4. APPLICATION

The purpose of this study is to identify the product groups that tend to sell together using the shopping records of a branch of a major hardware company. In the article, the Apriori and FP-Growth algorithm features of WEKA (Waikato Environment for Knowledge Analysis), an open-source data mining software, were used. SQL Server 2014 was used as a database.

4.1. Data Collection and Editing

The data used in this study belongs to the branch of a company which is one of the leaders of the domestic hardware market with more than 50 years of experience in the sector and an active business association with nearly twenty countries. In the study, company name will be kept confidential in accordance with the company's principle of confidentiality and the name will be listed as X Hardware, where necessary. X Hardware has 3081 different product types. The data set consisting of shopping movements between 21.08.2008 and 20.02.2014 has 81,384 different receipt information.

completed so that the noise in the data was cleared. In addition, the products were rearranged to be 13 basic categories and a total of 93 sub-categories. Main categories were numbered from 1 to 13, and subcategories belonging to each main category were represented by letters A, B, C and so on.

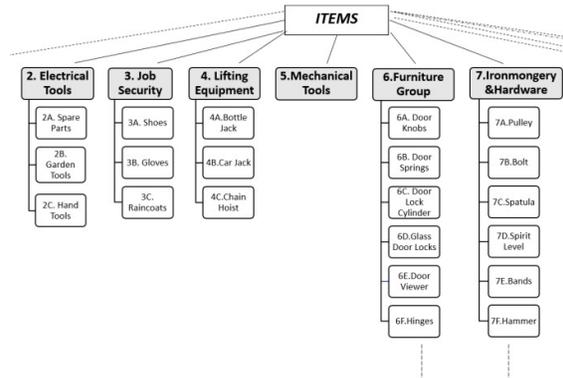


Figure 2.A section from the edited sub-product group table

Figure 1.The First Version of SQL Table

The table containing the receipts and product information in the receipts within the specified period was created by drawing SQL queries from the SQL database. Since our aim was to find products that were sold together within the receipts, the receipts containing only one product, regardless of the number of items, needed to be eliminated. With this new SQL query, the table was rearranged to keep the receipts containing at least two different products. At the end of the rearrangement, a total of 54,564 receipts containing at least two products remained. The table created in SQL was transferred to Excel as the receipt number, date of purchase, product, product group columns. When the table was first examined, it was seen that there were 293 different sub-product groups. However, it was observed that many of the product groups in the table were missing. By examining the data in detail, the deficiencies were

In order to compare the success of the extracted rules to the next periods, the data set containing information of successive years was divided into two at an appropriate rate. The data set containing records belonging to the first period of the data set was named 'Data Set-1', and the data set containing records belonging to the recent periods was named 'Data Set-2'. It is necessary to convert the data set including the date of purchase, receipt number, product name, product group columns into a form that WEKA can understand. Taking advantage of Excel's pivot table feature, it was converted to tables containing receipt numbers in rows and product categories in columns. Using simple Excel formulas and properties, these tables were filled with '1' if the receipt contained a product found in the column, and '0' if it did not. However, since the WEKA made rule deduction from a table composed of '0' and '1's: To illustrate, as it gives a result such as this: "Those who do not buy products in category 1A also do not buy products in category 5B.", 'T' was used instead of '1', and '?' was used instead of '0'. Through the conversion of the generated Excel file to 'arff' format, it was converted to a format that could be uploaded to the WEKA.

4.2. Running the data in the WEKA program

Using the Apriori and FP-Growth algorithms under the Associate option in the WEKA program, association rules were derived from the data set previously converted to 'arff' format. As a result of the experiments performed on a data

set at this size, the best rules were found to under 40% confidence and 0.5% support values.

First, the Apriori algorithm was run in the Weka program for Data Set-1. As a result of the algorithm with a running time of 13.05 minutes, it was seen that there were 24 rules which provide the minimum confidence and support values specified. The table below contains the top ten of these rules

Table 2.Results of Apriori Algorithm for Data Set-1

<i>Antecedent</i>	<i>Consequent</i>	<i>Confidence</i>	<i>Lift</i>
Chisels and Cutters, Bits Tip Adapters (677)	Drill Bits (431)	%64	3
Chisels and Cutters, Punch Group (366)	Drill Bits(233)	%64	3
Chisels and Cutters, Wire Brush (454)	Drill Bits (271)	%60	2.81
Chisels and Cutters (2370)	Drill Bits (1361)	%57	2.70
Meters, Nut Adapters (411)	Drill Bits (235)	%57	2.69
Diamond Saw, Chisels and Cutters (563)	Drill Bits (316)	%56	2.64
Meters, Nut Adapters (481)	Bits Tip Adapters (251)	%52	3.33
Nut Adapters, Drill Bits (962)	Bits Tip Adapters (497)	%52	3.29
Wire Brush, Nut Adapters (618)	Bits Tip Adapters (305)	%49	3.15
Bits Tip Adapters, Files(505)	Drill Bits (242)	%48	2.26

We can interpret the first rule of this table as follows: 431 of the 677 receipts containing {Chisels and Cutters, Bits Tips Adapters} group products also contain products from the {Drill Bits} group. In other words, customers who buy 'Chisel and Cutter' and 'Bits Tips Adapters' product have 64% likelihood of buying 'Drill Bits'. We can assess the lift values as the value of the interestingness of the rules. That is, the larger the lift value, the more interesting the rule is. For the first rule, we can interpret the lift value using the following formula;

$$\text{Percent increase} = (3 - 1) * 100 = 200\%$$

This can be interpreted as the fact that a receipt containing a product from the group of 'Chisels and Cutters' and 'Bits Tips Adapters' increases the likelihood of product inclusion by 200% from the 'Drill Bits' group [37]. Other rules in the table can be interpreted in the same manner.

The Apriori algorithm was then run under the same minimum trust and support values for Data Set-2, at the end of the algorithm that resulted in 41 seconds, it was seen that there were 19 rules that gave minimum confidence and minimum support values. The table below contains the top ten of these rules.

Table 3.Results of Apriori Algorithm for Data Set-2

<i>Antecedent</i>	<i>Consequent</i>	<i>Confidence</i>	<i>Lift</i>
Diamond Saw, Chisels and Cutters, (104)	Drill Bits (71)	%68	4.85
Chisels and Cutters, Punch Group (117)	Drill Bits (76)	%65	4.61
Chisels and Cutters, Bits Tip Adapters (155)	Drill Bits (99)	%64	4.54
Chisels and Cutters, (540)	Drill Bits (314)	%58	4.13
Punch Group, Nut Adapters (142)	Bits Tip Adapters (80)	%56	4.56
Nut Adapters, Drill Bits (206)	Bits Tip Adapters (111)	%54	4.36
SDS Plus Adapters (167)	Bits Tip Adapters (89)	%53	4.31
Wire Brush, Nut Adapters (146)	Bits Tip Adapters (77)	%53	4.27
Diamond Saw, Nut Adapters (128)	Bits Tip Adapters (66)	%52	4.17
Nut Adapters (637)	Bits Tip Adapters (313)	%49	3.98

Another algorithm that is used to determine association rules in the Weka program is the FP-Growth algorithm. The FP-Growth algorithm yielded results on Data Set-1 over 1 second while it gave results in less than 1 second on Data Set-2. The FP-Growth algorithm was run separately on Data Set-1 and Data Set-2, and the rules shown in the following tables were derived.

(637)	Adapters (313)		
-------	----------------	--	--

Table 4.FP-Growth Algorithm Results for Data Set-1

<i>Antecedent</i>	<i>Consequent</i>	<i>Confidence</i>	<i>Lift</i>
Coupling and Socket (721)	Spiral Air Hoses (329)	%46	15.39
Car Jack (1233)	Bottle Jack (509)	%41	4.64

In Weka, with the FP-Growth algorithm, two appropriate rules on Data Set-1 were derived. The first rule shows that out of the 721 receipts containing 'Coupling and Socket' group products, 329 receipts also contain products from the 'Spiral Air Hoses' group. The resulting rules show that the FP-Growth algorithm bypasses the rules with high confidence value when compared with the rules derived from the Apriori algorithm.

Table 5.FP-Growth Algorithm Results for Data Set-2

<i>Antecedent</i>	<i>Consequent</i>	<i>Confidence</i>	<i>Lift</i>
Diamond Saw, Chisels and Cutters, (104)	Drill Bits (71)	%68	4.85
Punch Group, Chisels and Cutters, (117)	Drill Bits(76)	%65	4.61
Bits Tip Adapters, Chisels and Cutters, (155)	Drill Bits (99)	%64	4.54
Chisels and Cutters, (540)	Drill Bits (314)	%58	4.13
Punch Group, Nut Adapters (142)	Bits Tip Adapters (80)	%56	4.56
Drill Bits, Nut Adapters (206)	Bits Tip Adapters (111)	%54	4.36
SDS Plus Adapters(167)	Bits Tip Adapters (89)	%53	4.31
Wire Brush, Nut Adapters (146)	Bits Tip Adapters (77)	%53	4.27
Diamond Saw, Nut Adapters (128)	Bits Tip Adapters (66)	%52	4.17
Nut Adapters	Bits Tip	%49	3.98

When the FP-Growth algorithm was run on Data Set-2, it was seen that the first ten rules were the same as the Apriori algorithm. However, when the running time of two algorithms is considered, it was seen that the FP-Growth algorithm yielded results much faster than the Apriori algorithm.

5. CONCLUSION

In this article, by using the Apriori and FP-Growth algorithms of the WEKA 3.8 data mining program, mining of association rules was realized on real data. Thus, it was aimed to identify product groups that tended to be sold together. As a result of the analysis carried out, it was seen in Table-1 and Table-2 that the strongest associations in the analysis period of the company were between {Chisels and Cutters}, {Bits Tips}, {Drill Bits}, {Punch Group}, {Meters}, {Screw Adapters} and {Files} The data set consisting of five and a half year sales data of X Hardware was divided into two data sets named Data Set-1 and Data Set-2.

First, Apriori and FP-Growth algorithms were applied separately on Data Set-1. When the algorithms were compared in terms of performance, it was seen that the FP-Growth algorithm yielded 781 times faster results than the Apriori in a data set at this size. However, the resulting rules showed that the FP-Growth algorithm failed to find the first 14 rules with high confidence value. Then Apriori and FP-Growth algorithms were applied for Data Set-2; it was again found that FP-Growth yielded results in less than 1 second, while Apriori yielded results in 41 seconds. However, it was seen that the number of rules derived from Data Set-2 and their order were the same in both algorithms.

Table-1 and Table-2 containing Apriori algorithm resultant rules of Data Set-1 and Data Set-2 were respectively compared. It was seen that the rules with important confidence values in Table-1 were also included in Table-2. For example; {Chisels and Cutters, Bits Tips Adapters} → {Drill Bits} is a rule (with 64% confidence value) drawn from the Data Set-1 which included the first period data. The fact that this rule was seen with same 64% confidence value in Data Set-2 which included consecutive period data can be interpreted that the rules derived from Data Set-1 are also largely valid in real life. The company was unaware of the rules derived in the early periods. Nevertheless, it was seen from the rules derived in the subsequent period that the sales were parallel. If the company had been aware of these associations, it could have increased its sales, in turn, its revenues by using these rules in catalog design, in-store

layout, warehouse design, discount- promotional arrangements and decision support system designs.

In the future studies, it is aimed to draw seasonal association rules by working on the relationship between the season and the hardware sector.

REFERENCES

- [1] G. Silaharoğlu, *Veri Madenciliği: Kavram ve Algoritmaları*, 3rd ed. İstanbul: Papatya Bilim, 2016.
- [2] G. G. Emel and Ç. Taşkın, "Genetik Algoritmalar ve Uygulama Alanları," *Uludağ J. Econ. Soc.*, vol. 21, no. 1, pp. 129–152, 2002.
- [3] D. Van Den Poel, J. De Schampelaere, and G. Wets, "Direct and indirect effects of retail promotions on sales and profits in the do-it-yourself market," *Expert Syst. Appl.*, vol. 27, no. 1, pp. 53–62, 2004.
- [4] Y. L. Chen, K. Tang, R. J. Shen, and Y. H. Hu, "Market basket analysis in a multiple store environment," *Decis. Support Syst.*, vol. 40, no. 2, pp. 339–354, 2005.
- [5] C. H. Yun, K. T. Chuang, and M. S. Chen, "Adherence clustering: An efficient method for mining market-basket clusters," *Inf. Syst.*, vol. 31, no. 3, pp. 170–186, 2006.
- [6] R. J. Kuo, S. Y. Lin, and C. W. Shih, "Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan," *Expert Syst. Appl.*, vol. 33, no. 3, pp. 794–808, 2007.
- [7] S. Erdem and G. Özdağoğlu, "Analyzing of Emergency Data Of A Training And Research Hospital In Aegean Region Using Data Mining," *Anadolu Univ. J. Sci. Technol.*, vol. 270, pp. 261–270, 2008.
- [8] E. Ergün, "Ürün Kategorileri Arasındaki Satış İlişkisinin Birliktelik Kuralları ve Kümeleme Analizi ile Belirlenmesi ve Perakende Sektöründe Bir Uygulama," Afyon Kocatepe University, 2008.
- [9] M. Timor and U. T. Şimşek, "Veri Madenciliğinde Sepet Analizi ile Tüketici Davranışı Modellemesi," *İstanbul Manag. J.*, vol. 59, pp. 3–10, 2008.
- [10] T. Li and X. Li, "Novel alarm correlation analysis system based on association rules mining in telecommunication networks," *Inf. Sci. (Ny)*, vol. 180, no. 16, pp. 2960–2978, 2010.
- [11] W. Y. Chiang, "To mine association rules of customer values via a data mining procedure with improved model: An empirical case study," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1716–1722, 2011.
- [12] J. Nahar, K. S. Tickle, A. B. M. S. Ali, and Y. P. P. Chen, "Significant cancer prevention factor extraction: An association rule discovery approach," *J. Med. Syst.*, vol. 35, no. 3, pp. 353–367, 2011.
- [13] M. Durdu, "Application of Data Mining in Customer Relationship Management Market Basket Analysis in a Retailer Application of Data Mining in Customer Relationship Management Market Basket Analysis in a Retailer Store," Dokuz Eylül University Graduate, 2012.
- [14] S. Erpolat, "Comparison of Apriori and FP-Growth Algorithms on Determination of Association Rules in Authorized Automobile Service Centres," *Anadolu Univ. J. Soc. Sci.*, vol. 12, no. 2, pp. 137–146, 2012.
- [15] I. Cil, "Consumption universes based supermarket layout through association rule mining and multidimensional scaling," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 8611–8625, 2012.
- [16] L. Hanguang and N. Yu, "Intrusion Detection Technology Research Based on Apriori Algorithm," *Phys. Procedia*, vol. 24, pp. 1615–1620, 2012.
- [17] E. Güngör, N. Yalçın, and N. Yurtay, "Apriori Algoritması ile Teknik Seçmeli Ders Seçim Analizi," in *International Congress on Education, Distance Education and Educational Technology*, 2013.
- [18] A. Mansur and T. Kuncoro, "Product Inventory Predictions at Small Medium Enterprise Using Market Basket Analysis Approach-Neural Networks," *Procedia Econ. Financ.*, vol. 4, no. Icsmed, pp. 312–320, 2012.
- [19] P. Paranjape-Voditel and U. Deshpande, "A stock market portfolio recommender system based on association rule mining," *Appl. Soft Comput. J.*, vol. 13, no. 2, pp. 1055–1063, 2013.
- [20] M. Dhanabhakya and M. Punithavalli, "An Efficient Market Basket Analysis based on Adaptive Association Rule Mining with Faster Rule Generation Algorithm," *SIJ Trans. Comput. Sci. Eng. its Appl.*, vol. 1, no. 3, pp. 105–110, 2013.
- [21] J. Nahar, T. Imam, K. S. Tickle, and Y. P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females,"

- Expert Syst. Appl.*, vol. 40, no. 4, pp. 1086–1093, 2013.
- [22] Z. Ceylan, “Shelf Arrangement Via Data Mining Techniques and A Case Study Shelf Arrangement Via Data Mining,” Marmara University, 2014.
- [23] B. Doğan, B. Erol, and A. Buldu, “Sigortacılık Sektöründe Müşteri İlişkileri Yönetimi İçin Birlikte Kullanılması,” *Marmara Univ. J. Sci.*, vol. 26, no. 3, p. 128, 2014.
- [24] S. Kamley, S. Jaloree, and R. S. Thakur, “An Association Rule Mining Model for Finding the Interesting Patterns in Stock Market Dataset,” *Int. J. Comput. Appl.*, vol. 93, no. 9, pp. 975–8887, 2014.
- [25] M. Kaur and S. Kang, “Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining,” *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 78–85, 2016.
- [26] D. Solnet, Y. Boztug, and S. Dolnicar, “An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue,” *Int. J. Hosp. Manag.*, vol. 56, pp. 119–125, 2016.
- [27] P. Yazgan and A. O. Kusacki, “A Literature Survey on Association Rule Mining Algorithms,” *Southeast Eur. J. Soft Comput.*, vol. 5, no. 1, pp. 5–14, 2016.
- [28] M. Özçalıcı, “Veri Madenciliğinde Birlikte Kullanılması ve İkinci El Otomobil Piyasası Üzerine Bir Uygulama,” *Ordu Univ. J. Soc. Sci. Res.*, vol. 7, no. March, pp. 45–58, 2017.
- [29] P. Yildirim, D. Birant, and T. Alpyildiz, “Discovering the relationships between yarn and fabric properties using association rule mining,” *Turkish J. Electr. Eng. Comput. Sci.*, vol. 25, pp. 4788–4804, 2017.
- [30] P. Yildirim, D. Birant, and T. Alpyildiz, “Discovering the relationships between yarn and fabric properties using association rule mining,” pp. 1–17, 2017.
- [31] W. J. Frawley, G. Piatetsky-shapiro, and C. J. Matheus, “Knowledge Discovery in Databases: An Overview,” *AI Mag.*, vol. 13, no. 3, pp. 57–70, 1992.
- [32] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. 2006.
- [33] A. S. Koyuncugil and N. Özgülbaş, “Veri Madenciliği :Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları,” *Int. J. Informatics Technol.*, vol. 2, no. 2, pp. 21–32, 2009.
- [34] T. C. Yang and H. Lai, “Comparison of product bundling strategies on different online shopping behaviors,” *Electron. Commer. Res. Appl.*, vol. 5, no. 4, pp. 295–304, 2006.
- [35] K. J. Cios, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan, *Data Mining: A Knowledge Discovery Approach*, 1st ed. Springer, 2007.
- [36] E. Alpaydın, “Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri,” in *Bilişim 2000 Eğitim Semineri*, 2000.
- [37] “Association Analysis Simplified,” *Analytics Trainings*, 2018. [Online]. Available: <http://analyticstrainings.com/?cat=5>.
- [38] R. Agrawal, T. Imielinski, and A. Swami, “Mining Association in Large Databases,” *Proc. 1993 ACM SIGMOD Int. Conf. Manag. data - SIGMOD '93*, pp. 207–216, 1993.
- [39] R. Srikant and R. Agrawal, “Mining generalized association rules,” *Futur. Gener. Comput. Syst.*, vol. 13, no. 2–3, pp. 161–180, 1994.
- [40] F. C. Özçakır and A. Y. Çamurcu, “Birlikte Kullanılması İçin Bir Veri Madenciliği Yazılımı Tasarımı ve Uygulaması,” *Istanbul Commer. University J. Sci.*, vol. 12, pp. 21–37, 2007.
- [41] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” *94 Proc. 20th Int. Conf. Very Large Data Bases*, vol. 1215, pp. 487–499, 1994.
- [42] A. Döşlü, “Veri Madenciliğinde Market Sepet Analizi ve Birlikte Kullanılmasının Belirlenmesi,” Yıldız Technical University, 2008.
- [43] W. Zhang, H. Liao, and N. Zhao, “Research on the FP growth algorithm about association rule mining,” *2008 Int. Semin. Bus. Inf. Manag. ISBIM 2008*, vol. 1, pp. 315–318, 2009.