Measuring Customer Purchasing Behavior based on Apriori Algorithm for Creating an Efficient Marketing Strategy

Murat AKCIL[1] ORCID *0000-0003-4963-1826*
Serap AKCAN[*2] ORCID *0000-0003-2621-9142*

[1]*Süleyman Demirel University, Faculty of Engineering, Department of Industrial Engineering, Isparta, Türkiye*
[2]*Tarsus University, Faculty of Engineering, Department of Industrial Engineering, Mersin, Türkiye*

Atıf şekli/ How to cite: AKCIL, M., AKCAN, S., (2023). Measuring Customer Purchasing Behavior based on Apriori Algorithm for Creating an Efficient Marketing Strategy. Cukurova University, Journal of the Faculty of Engineering, 38(3), 751-764.

Abstract

Due to technological developments and competitive conditions, it is becoming more and more important to gain customers and keep them. Association rules are one of the methods used effectively to determine which products and services are more important and preferred by customers together. The aim of this study is to determine the types of patterns purchased by a company operating in the forest products sector, using the data of decor paper, which affects the cost the most, is used extensively in production and is difficult to plan. This study was carried out with the Apriori algorithm in WEKA, which is widely used in association analysis. It is thought that the association rules obtained from this study will enable the firm to determine the purchasing behaviors of its customers by revealing the overseas model trends and associations. In addition, this study will shed light on the company at the stage of determining the marketing strategy of the company.

Keywords: Big data, Data mining, Association rules, Apriori algorithm, Market basket analysis

Etkin Bir Pazarlama Stratejisi Oluşturmak için Apriori Algoritması Temelinde Müşteri Satın Alma Davranışının Ölçülmesi

Öz

Teknolojik gelişmeler ve rekabet koşulları nedeniyle müşteri kazanmak ve elde tutmak giderek daha önemli hale gelmektedir. Birliktelik kuralları, hangi ürün ve hizmetlerin müşteriler tarafından birlikte daha önemli ve tercih edildiğinin belirlenmesi amacıyla etkin olarak kullanılan yöntemlerden biridir. Bu çalışmanın amacı, orman ürünleri sektöründe faaliyet gösteren bir firmanın, maliyeti en çok etkileyen, üretimde yoğun olarak kullanılan ve planlanması zor olan dekor kağıdı verilerini kullanarak, satın aldığı desen türlerinin belirlenmesidir. Bu çalışma, birliktelik analizinde yaygın olarak kullanılan WEKA'daki Apriori algoritması ile gerçekleştirilmiştir. Bu çalışmadan elde edilen birliktelik kurallarının firmanın yurt dışı model trendlerini ve birlikteliklerini ortaya çıkararak müşterilerinin satın alma davranışlarını belirlemesine olanak

---

[*]Sorumlu yazar (Corresponding Author): Serap AKCAN, *serapakcan@tarsus.edu.tr*

sağlayacağı düşünülmektedir. Ayrıca bu çalışma firmanın pazarlama stratejisini belirleme aşamasında firmaya ışık tutacaktır.

Anahtar Kelimeler: Büyük veri, Veri madenciliği, Birliktelik kuralları, Apriori algoritması, Pazar sepeti analizi

1. INTRODUCTION

Increasing competition conditions and the ease of instantly comparing the products demanded by customers, thanks to technological developments, make it increasingly difficult for companies to gain customers and retain existing customers. Marketing activities have become very important for companies that want to survive in these conditions. The right marketing activities should be managed with the right marketing plan. It is important that all activities, from promotions to campaigns, should be in plan to gain new customers, to satisfy existing customers and to increase awareness. But these activities mean costs for the firm. Activities that are not correctly positioned and analyzed will return to the company as an extra cost rather than gain. Marketing strategies determined only by experience and knowledge are not sufficient in a period when technology is so prominent and competition is so intense. Companies need to use technology and data science in order to stay one step ahead of their competitors and make a difference.

Today, almost all companies use an ERP software. ERP software; they are important structures used in all departments of companies from accounting to purchasing, from production to planning and connecting all units and processes [1]. Thanks to these systems, records of all works can be kept. But this unprocessed and raw data is just a chunk of data unless properly analyzed and made sense of. This is where the concepts of data mining and big data come into play. Data mining is making a large number of data interpretable in line with desired targets or categorizing the desired data from big data [2].

Association rule mining is one of the analysis methods used to detect purchasing tendencies and behaviors of customers. If the rules formed after the analysis are interpreted correctly, it can be a very important guide when determining marketing activities. The knowledge of which products or services customers give more importance to or prefer together can also be beneficial in using the budget spent for marketing activities more effectively.

In the literature, it is possible to come across many studies on association analysis and market basket analysis in different fields. Angeline [3] focused on predicting the performance of students by analyzing data such as homework, internal evaluation tests, participation performance in an educational institution's database with the Apriori algorithm. In the analyzes made, the students were classified as good, average and below average. With this system, it is thought that education can be given and directed according to the level of the students. Qiao et. al. [4] used association rules to analyze the correlation between the prices of 6 types of agricultural products they selected in 2014. They concluded that there was a strong correlation between the price of corn and the price of soybeans in 2014. It is thought that the study will play a positive role in price analysis and estimation of agricultural products. Guo et. al. [5] proposed a new method called MWLR for wind power estimation, which is very important for wind energy investments, which are at the forefront of renewable energies and whose use is increasing. This method generally consists of 4 parts. These are clustering using the K-Means approach, using the Apriori algorithm to discover association rules, estimating the wind speed according to the chaotic time series prediction model, and correcting the predicted wind speed data using previously discovered related rules. The results obtained as a result of the simulations show that the MWLR method can predict the wind speed effectively. Zulfikar et. al. [6] conducted an association analysis by using the sales history of more than 10,000,000 records of a retail company named XMART, which has a portfolio of more than 5,500 products, as a data set. While performing the association analysis with the

Apriori algorithm, the aim is to make the promotions of the company more effective and to provide inputs that will help to prepare a layout plan using the rules. Xi et. al. [7] deduced that the classification should include all four basic traffic parameters as environment, vehicle, road and driver while taking into account the causal factors of traffic accidents in their studies. In the study, the most important causal factors were ranked in order of importance using the Analytical Hierarchy Process (AHP) technique, and the degree of accident and the level of impact were analyzed with the Apriori association algorithm. In this way, it is aimed to determine the type and severity of accidents. Sevri et. al. [8] aimed to find the associations between the qualities of independent criminal records by using the Apriori algorithm. It is expected that the rules of association that emerged in the study conducted using the criminal record data recorded in the USA in 2013 will be useful for crime analysis. Considering the rules formed between different criminal record qualifications, it is possible to establish a relationship between new and old events. Mirmozaffari et. al. [9] with the knowledge that approximately 60% of the world population has a heart disease, using a clinical data set they performed analyzes with the Apriori algorithm and claiming that in the diagnosis of coronary heart diseasesassociation analysis can be used to extract. The data set consists of 209 samples and 8 features and it is thought that the study will be a light for the relevant doctors. Cho and Kim [10] aimed to measure the text similarity between two texts based on classifications algorithms. Beheshtian-Ardakani et. al. [11] proposes a new model for product classification and grouping in websites based on e-commerce with market segmentation variables and loyalty analysis. Customers were first grouped by measuring their loyalty with the RFM (recency, frequency, and monetary value) model, and each group was divided according to the market segmentation variable. RFM Analysis is an analysis method performed by using transaction history information according to how recently (recency), how often customers shop (frequency) and how much they spend (monetary value). By using the Apriori algorithm, the rules for which products are sold together are determined and the products are

divided into certain groups. Then, which group of products would be recommended to which group of customers was determined by classification models. The model proposed in this study increases the chance of success as the target customers are determined in direct marketing and adopts the idea that the sales chance will be increased because it recommends the product groups to the right customer. Zainol et. al. [12] aimed to detect cyberbullying cases by using the Apriori algorithm, one of the association analysis algorithms in their studies. Both the high level of trust and the excess of rules at this level have revealed that association rules are effective in detecting cyberbullying cases. Kurnia et. al. [13] analyzed the sales data they received from a fish restaurant and investigated which of the products on the menu customers prefer to buy together with the association analysis method. It aimed to help the company create a potential promotional strategy with certain product combinations by interpreting user preferences with the established association rules. Hidayat et. al. [14], in this study, it is focused on the determination of which products are associated with the behavior of Breiliant cosmetics store customers by using the market basket analysis method. Analysis was performed using Apriori and FP-Growth algorithms. Monthly sales transaction data of materials in the Breiliant Store in November 2018 were used for the analyses. The most striking of the results is the rule that Original Liquid Bleaching Seeds, Harva Peeling Gel and Castor oil should be sold together at the 30% confidence level. Simanjorang and Sijabat [15] conducted an analysis of customers' printing needs using sales data of a printing company. In the analysis study using the Apriori algorithm, it was tried to determine which products were purchased or ordered together. The rule has emerged that those who order cake boxes at a support value of 55% and a confidence interval of 90% are placing orders in the brochure. Dorling et. al. [16] identified the genes most useful for predicting breast cancer risk in their study and used association analysis to provide estimates of risks associated with protein truncation variants to guide genetic counseling. Edastama et. al. [17] mentioned in their studies that analyzing the sales data of companies correctly will be beneficial to improve sales and to deliver more effective product sets

when determining promotional products. In the study, which was conducted using the sales data of a national optics company, association analysis was used to determine the most popular and demanded products as well as related products. It is aimed to create useful information while determining the sales and marketing activities of the company with the rules formed as a result of the analyzes made with the Apriori algorithm. Pan [18] performed association analysis using the Apriori algorithm to evaluate students' physical fitness based on physical education data and to reveal hidden correlations between physical fitness indices. It is thought that the resulting rules provide an important guide for university students for physical education test and curriculum designs. Ünvan [19] performed grocery basket analysis with the sales data of any supermarket on the Vancouver Island University website. Apriori and FP Growth algorithms were used to analyze association rules on the data set containing 225 different products, but the Apriori algorithm did not yield results because the data set was multi-category. The first 10 rules given by the FP Growth algorithm are interpreted. The product placement study is aimed by using the established rules. Zhao and Keikhosrokiani [20] proposed a combination of various algorithms for sales forecasting and product recommendation analysis, considering the changing conditions during the pandemic period. He used the RFM analysis method to classify customers according to their level, XGBoost and Random Forest algorithms for purchase forecasting, and association rules theory and Apriori algorithm for product forecasting. They stated that this combination will provide preliminary data to companies on both the purchasing and marketing sides. Mirhashemi and Mirzaei [21] used unified clustering algorithms and association rules for better management of the amount of water supplied to the Abyek Plain irrigation network in Iran in their study. The study, which uses K-Means algorithm for combined clustering and Apriori algorithm for association analysis, aims to provide a different perspective to prevent water wastage since there is water shortage and lack of precipitation in the Abyek plain.

Looking at the literature, no study has been found in the forest products sector using the Apriori

algorithm. Therefore, in this study, association rules were tried to be determined by using real data obtained from a company operating in the forest products sector. For this purpose, decor papers with 88 patterns in the current color chart, which are difficult to plan due to the diversity of the raw materials purchased by the company, are selected for association analysis. Considering these decors, the company's monthly and annual melamine coated particle board (MCPB) orders between 2009-2021 were determined and association analyzes were carried out.

2. METHODOLOGY

The flow presented for the method and solution used in the study is shown in Figure 1.

After the data is collected through SAP, the sales data are exported. Then, the data are selected and preprocessed according to the analysis process. After this, the data is cleaned, filtered and transformed in WEKA program so that the association analysis can be done correctly. Finally, the ready data set is analyzed with the help of WEKA program and association rules are formed.

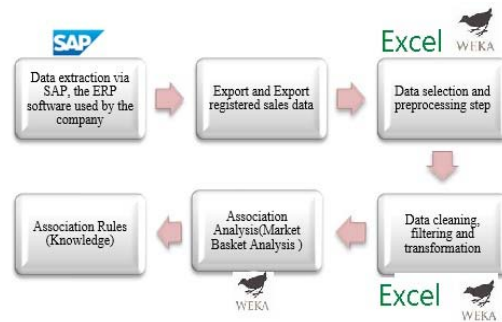**Figure 1.** Flow chart for methodology

2.1. Association Rules

Many methods are used to detect purchasing tendencies and decision-making behaviors of customers. One of them is association analysis. Association analyzes are analyzes that allow us to see the strong or weak links in data sets around certain rules. It is revealed how there is a connection between the association analysis and the events that

take place. The analyst sets a success rate when launching the app, and rules equal to or greater than this rate are determined by the app.

Association analysis is especially important for companies' sales development strategies. The rules obtained from this analysis constitute information for the purchasing habits of the companies, for choices such as which products will be placed in catalogs or how they will be placed on the shelves. Association analysis is also used in web page clicks and in many other areas [22]. Because, while calculating the basket analysis, the connection between goods and services, support and trust criteria are taken as basis. Association Rules, calculated according to the level of support and reliability, indicate the potential relationships of the data. A strong Association Rule has great support and a high level of credibility.

2.2. Market Basket Analysis

It is a type of analysis used to determine the association and bond between the products purchased by the customers in their shopping. It gives important information to the companies about the established association rules and the purchasing habits of the customers. In the essence of this information, the information about which product is purchased with which product is learned. Market basket analysis can be briefly defined as a technique to find consumer habits from datasets [23]. In market basket analysis, large datasets take into account the purchasing patterns of users. Association rules are the rules that emerge from the analysis and are of great importance for finding frequently purchased items and purchasing habits. If item A is bought, then item B is bought as well [24]. Market basket analysis, which is especially used to determine marketing strategies, is to determine the relationships according to the analysis of the data formed according to the consumption behaviors of the customers with the association rules techniques.

2.3. Apriori Algorithm

Algorithms, such as Apriori, Eclat and Fp-Growth, are used to search for association rules. Agrawal et

al published the Apriori algorithm in 1993, making a significant improvement in the extraction of association rules. The Apriori algorithm gets its name from the fact that it takes the information from the previous step. With its publication, it has become one of the most used association rules algorithms [24].

The Apriori algorithm allows to use support and minimum confidence parameters in order to find useful, more accurate and desired rules from the data set or related database [25]. In the analysis, the minimum support value and the confidence threshold are specified, and the rules above these parameters are displayed and analyzed. Thus, it allows for more flexible and desired analyzes in large datasets or, on the contrary, smaller datasets.

Finding repetitive datasets from big data is one of the jobs of data mining. The Apriori algorithm, on the other hand, can not only find repetitive data sets, but also includes topics such as hierarchical relationship rules, sequential pattern creation, and inferring meaning from the section. The Apriori algorithm can be defined in three steps to create rules from repetitive data sets. The first step is to determine a path for the keywords, the second step is to define the data structures used, and the last step is to turn it into code with an algorithm [26].

Apriori creates candidate item sets by combining large item sets obtained in the previous cycle and deleting subsets obtained as small item sets in the previous cycle regardless of transactions in the database. The solution form of the Apriori cycle, on the other hand, presents an iterative solution proposal. The underlying object sets are realized by growing by incorporating a subset of them. As a first step, the joining criteria are defined in order to create the first layered object sets, the frequently used elements in the transitions in the subset are preferred to cluster by creating the candidate of the object it passes through. Participation criteria are defined by making important examinations in the data set, so the processes are repeated sequentially until they decrease [24].

The rules that are reached after the analysis of the Apriori algorithm, namely associations, are

expressed with certain parameters. The most important of these parameters can be seen as support and trust. While the support values express the frequency of association of the items in the rule, the confidence value shows the percentage with which the rule is correct. Let's assume that the X and Y expressions refer to two different sets of materials. As seen in equation (1), while calculating the support value for X, the number of material clusters X is divided by the total number of material clusters. In equation (2), the support value of the material clusters X and Y and the probability of coexistence in all clusters are calculated. The probability that the Y material cluster will be found in the X material cluster is called the confidence value and is calculated by equation (3) or equation (4).

$$Support\ (X) = \frac{X\ number\ of\ material\ clusters}{Total\ number\ of\ material\ clusters} \qquad (1)$$

$$Support\ (X,Y) = \frac{(X,Y)\ number\ of\ material\ clusters}{Total\ number\ of\ material\ clusters} \qquad (2)$$

$$Confidence\ (X,Y) = \frac{(X,Y)\ number\ of\ material\ clusters}{X\ number\ of\ material\ clusters} \qquad (3)$$

$$Confidence\ (X,Y) = \frac{Support\ (X,Y)}{Support\ (X)} \qquad (4)$$

These values allow us to see the accuracy and reliability of the rules formed in the association analysis. The higher these values, the more important the association. While performing the analysis, first of all, the support and trust values are determined and the rules within these values are obtained.

3. IMPLEMENTATION

The forest products sector, as the name suggests, includes the production and trade of various goods that take their raw materials from the forest and generally using wood. Timber, parquet, particle board, fiberboard, paper, cardboard, logs are the main forest products.

A Turkish forest products industry, with its development in recent years, is among Turkey's fastest growing sectors. In addition, the foreign trade balance of many sub-sectors, especially the furniture and wooden plate sector, is progressing positively. Turkey ranks 13th among 166 countries in the world with a share of 1.6% in timber production. Turkey's share in world furniture production is around 1% and it is increasing rapidly. Some organizations in Turkey, especially in the wood panel sector, have become one of the world's leading companies. About 2.4% of the added value created in the Turkish manufacturing industry belongs to the forest products sector. Türkiye has established export and import connections with more than 200 countries in the field of forest products. It is seen that a tendency to open up to new markets in the sector. New technology and business applications are becoming widespread [27].

The data used in this study were taken from a company operating in the plate industry in Turkey. The plate industry can be examined in terms of two different products. The first is fiberboard. Fiberboard, known in the industry as MDF, is a type of board in homogeneous condition, formed by using lignocellulosic fibers with felting and naturally structuring properties. It is produced as a result of drying or pressing the board, especially with the use of additional glue of fiber bundles. The second is particleboard. Particle Board is a material that is formed by turning chips or small particles obtained from lignocellulosic raw materials into large and large-surfaced boards under heat and pressure with the help of a synthetic resin or a suitable adhesive, and is used especially in the furniture industry [28].

Apart from these, there is also a Melamine Coated Particle Board (MCPB). MCPB is wood-based decorative plates that are produced as a result of pressing the impregnated decor paper and the particleboard under temperature and pressure for a certain period of time. Decor papers with various colors and patterns are the most important material that makes these plates decorative.
The data to be used for the association analysis were created by pulling the company's domestic, export and export registered sales data between the years 2009-2021 through SAP annually and then combining them. The main reason for pulling the
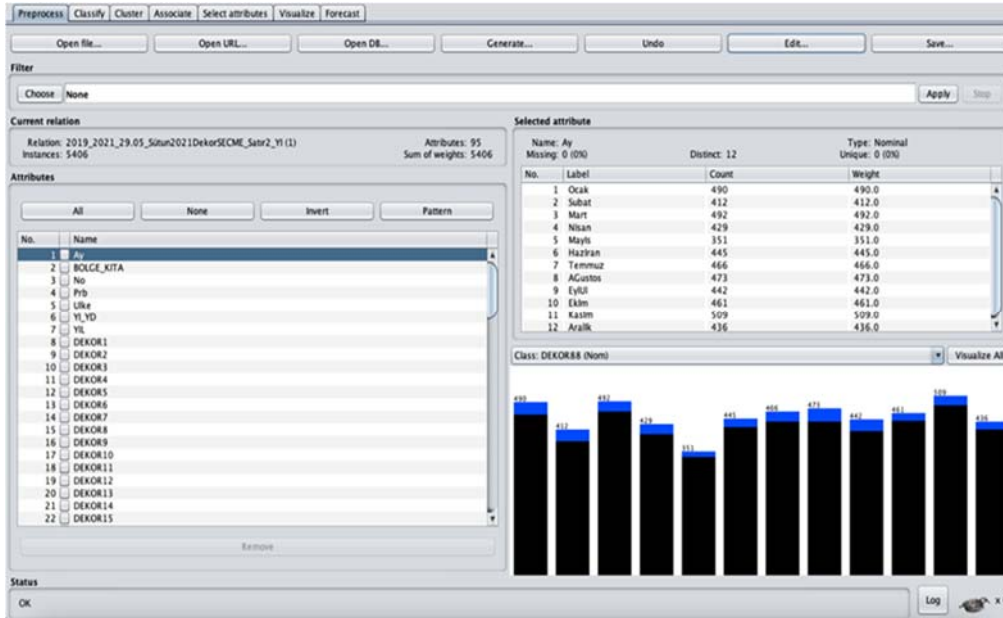
data annually is the system inconvenience due to the large data being transferred from SAP to Excel.

It is seen that the company sold 956 different designs of MCPB between 2009-2021. When the data is combined with domestic, export and export records, it includes 55863 different orders. However, the company does not ensure the continuity of some patterns according to current trends and stops selling the pattern. Since the stock replenishment of the discontinued decors is not continued, not being considered in the study will allow the association rules to provide more useful information. Therefore, the number of patterns was chosen to be 88. In addition, the data set is divided into foreign and domestic, due to the fact that foreign and domestic sales trends are different from each other. In this study, foreign data sets were used.

In this study, data sets were prepared in two different ways in order to see how the rules have changed compared to the past, to take into account that customers are constantly changing, and to examine the effect of the Covid 19 pandemic. Data sets were determined as foreign sales between 2009-2021 (Data set-1) and foreign sales between 2019-2021 (Data set-2).

There is country information for foreign sales in the data obtained from the SAP system. Since regional associations are also considered in the analysis to be made, the country information in the foreign sales data has been changed to continent. Datasets were updated in the light of this information; deficiencies were scanned in detail and noisy data were cleaned. Since the purpose of association analysis is to find patterns that are sold together, single-pattern orders are deleted. In order to convert the created data sets to "arff" format, the data set was loaded into the application using the Experimenter tab of the WEKA program (Figure 2). After these processes, the Open Explorer function was run and the Explorer module was switched (Figure 3).

By using the "NumericToNominal" feature under the Filters tab in the Explorer module, all of the attributes are converted to nominal format in order to make the data suitable for analysis with the Apriori algorithm. All of these processes were applied to the data sets and the data sets were saved in the "arff" format and made ready for analysis. As can be seen in Figure 4, the analysis of the prepared data sets was made under the 10% confidence and 0.5% support values as a result of the trials. The confidence interval value was entered as "0.1" in the "minMetric" section and the support value as "0.05" in the "lowerBoundMinSupport" section. The reason for keeping the confidence value low is the desire to see not the first 10 rules but also the following all rules.

Figure 2. WEKA experimenter module analyze tab

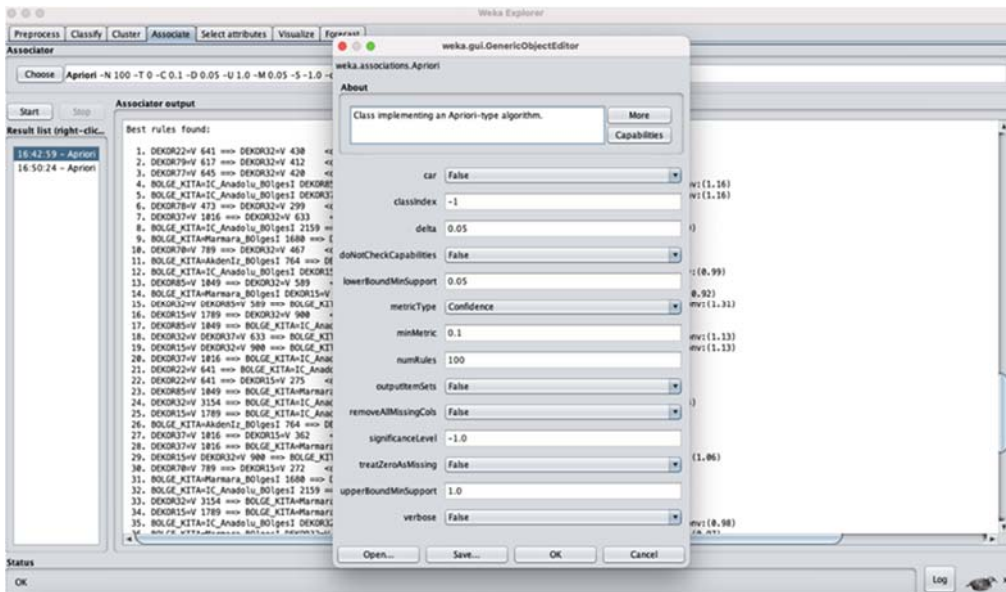**Figure 3.** WEKA explorer module preprocess tab

Figure 4. WEKA explorer module associate page

4. COMPUTATIONAL RESULTS AND DISCUSSION

In Figure 5, the rules formed as a result of the analysis of foreign sales data for the years 2009-

2021, which is called Dataset -1, using the Apriori algorithm are listed.

The 25 rules, provide support and minimum confidence intervals, have been displayed as output

before starting the analysis. Although there are more than 25 rules, the first 10 rules are transferred to Table 1 to examine in detail.

As seen in Table 1, DECOR84 is seen as the intersection point of associations. When we analyze the first rule, we see that out of 278 deliveries containing the DECOR19 and DECOR49 patterns, 211 were found in the DECOR84 pattern. In other words, customers who buy the DECOR19 and DECOR49 pattern together also buy the DECOR84 pattern at a rate of 76%. A value of 76% indicates the confidence value. In addition, the support value is seen as 3%. It shows that in 3% of all orders these three decors are purchased together. The lift value of 1.87 can be called the difference or interestingness value. The magnitude of the lift value shows how interesting the rule is. The percentage effect of the lift value and how much the sales between products affect the sales between each other can also be calculated. The remaining 9 rules can be explained as follows;

- Customers who buy DECOR 34 and DECOR59 patterns also buy DECOR84 pattern with 75% probability. In 0.03 of all deliveries, these products are sold together. The lift value is 1.86.
- Customers who buy DECOR49 and DECOR86 patterns also buy DECOR84 pattern with 74% probability. In 0.02 of all deliveries, these products are sold together. The lift value is 1.83.
- Customers who buy DECOR50 and DECOR59 patterns also buy DECOR84 pattern with 74% probability. In 0.03 of all deliveries, these products are sold together. The lift value is 1.82.

- Customers who buy DECOR19 and DECOR72 patterns also buy DECOR84 pattern with 73% probability. In 0.02 of all deliveries, these products are sold together. The lift value is 1.79.
- Customers who buy DECOR19 and DECOR59 patterns also buy DECOR84 with 71% probability. In 0.03 of all deliveries, these products are sold together. The lift value is 1.74.
- Customers who buy DECOR18 and DECOR72 patterns also buy DECOR84 pattern with 70% probability. In 0.02 of all deliveries, these products are sold together. The lift value is 1.73.
- Customers who buy DECOR18 and DECOR59 patterns also buy DECOR84 pattern with 69% probability. In 0.02 of all deliveries, these products are sold together. Lift value is 1.70.
- Customers who buy DECOR49 and DECOR72 patterns also buy DECOR84 pattern with 69% probability. In 0.03 of all deliveries, these products are sold together. Lift value is 1.70.
- Customers who buy DECOR49 and DECOR70 patterns also buy DECOR84 pattern with 68% probability. In 0.03 of all deliveries, these products are sold together. The lift value is 1.69.

In Figure 6, the rules formed as a result of analyzing the international sales data between 2019-2021, called Dataset - 2, using the Apriori algorithm are listed.

The 25 rules, provide support and minimum confidence intervals, have been displayed as output before starting the analysis. Although there are more than 25 rules, the first 10 rules are transferred to Table 2 to examine in detail.

Table 1. Apriori algorithm results for dataset-1 (the first 10 rules)

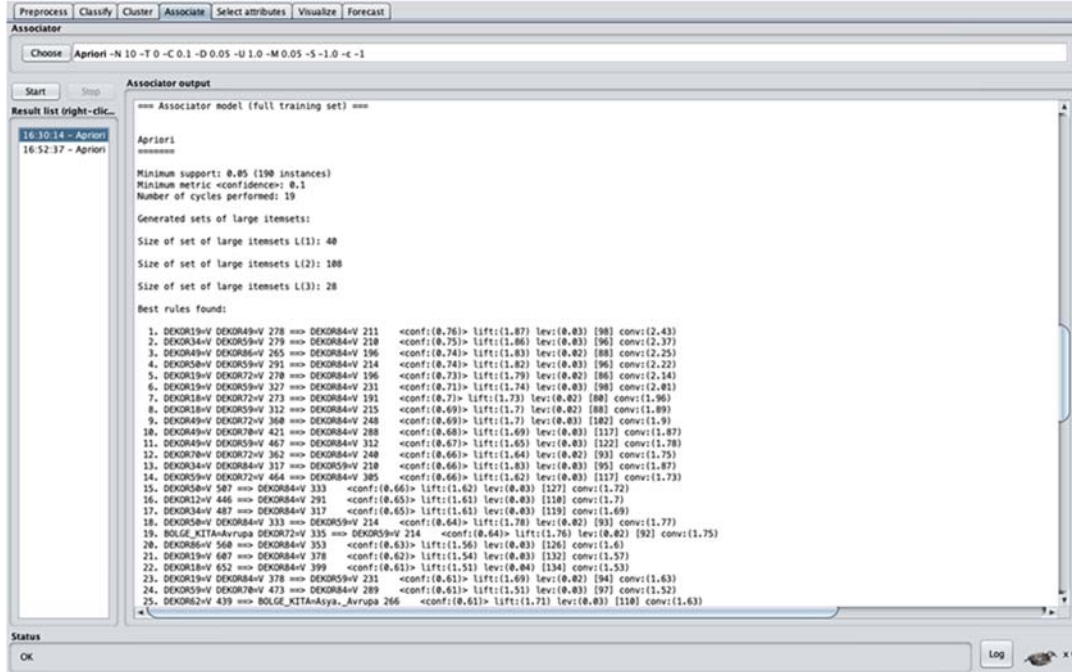| Premise	Results	Confidence	Support	Lift
DECOR19 & DECOR49 (278)	DECOR84 (211)	76%	0,03	1,87
DECOR34 & DECOR59 (279)	DECOR84 (210)	75%	0,03	1,86
DECOR49 & DECOR86 (265)	DECOR84 (196)	74%	0,02	1,83
DECOR50 & DECOR59 (291)	DECOR84 (214)	74%	0,03	1,82
DECOR19 & DECOR72 (270)	DECOR84 (196)	73%	0,02	1,79
DECOR19 & DECOR59 (327)	DECOR84 (231)	71%	0,03	1,74
DECOR18 & DECOR72 (273)	DECOR84 (191)	70%	0,02	1,73
DECOR18 & DECOR59 (312)	DECOR84 (215)	69%	0,02	1,7
DECOR49 & DECOR72 (360)	DECOR84 (248)	69%	0,03	1,7
DECOR49 & DECOR70 (421)	DECOR84 (288)	68%	0,03	1,69

Figure 5. Apriori analysis results for dataset –1

Table 2. Apriori algorithm results for dataset-2 (the first 10 rules)

Premise	Results	Confidence	Support	Lift
DECOR32 & DECOR62 (101)	Free Zone (87)	86%	0,03	1,74
DECOR54 (74)	Free Zone (63)	85%	0,02	1,72
Europe & DECOR59 (87)	DECOR32 (70)	80%	0,02	1,49
DECOR22 & DECOR62 (92)	Free Zone (70)	76%	0,02	1,54
DECOR62 (190)	Free Zone (143)	75%	0,05	1,52
DECOR32 & DECOR85 (74)	Free Zone (55)	74%	0,02	1,5
Europe (220)	DECOR32 (157)	71%	0,04	1,32
DECOR58(80)	Free Zone (56)	70%	0,02	1,41
Free Zone & DECOR85 (80)	DECOR32 (55)	69%	0,01	1,28
DECOR56(81)	DECOR32 (55)	68%	0,01	1,26

The first 10 rules in Table 2 can be explained as follows;

- Customers who buy DECOR32 and DECOR62 patterns are in the Free Zone with a probability of 86%. This rule applies to 0.03 of all deliveries. The lift value is 1.74.
- Customers who buy DECOR54 pattern are in the Free Zone with a probability of 85%. This rule applies to 0.02 of all deliveries. The lift value is 1.72.

- Customers in Europe who buy DECOR59 pattern also buy DECOR32 pattern with 80% probability. This rule applies to 0.02 of all deliveries. The lift value is 1.49.
- Customers who buy DECOR22 and DECOR62 patterns are in the Free Zone with a 76% probability. This rule applies to 0.02 of all deliveries. The lift value is 1.54.
- Customers who buy DECOR62 pattern are in the Free Zone with a 75% probability. This rule applies to 0.05 of all deliveries. The lift value is 1.52.

- Customers who buy DECOR32 and DECOR85 patterns are in the Free Zone with a probability of 74%. This rule applies to 0.02 of all deliveries. Lift value is 1.50.
- Customers in Europe also buy from DECOR32 pattern with a probability of 71%. This rule applies to 0.04 of all deliveries. The lift value is 1.32.
- Customers who buy the DECOR58 pattern are in the Free Zone with a 70% probability. This

rule applies to 0.02 of all deliveries. The lift value is 1.41.
- Customers in the Free Zone who buy DECOR85 pattern also buy DECOR32 pattern with a probability of 69%. This rule applies to 0.01 of all deliveries. The lift value is 1.28.
- Customers who buy DECOR56 pattern also buy DECOR32 pattern with 68% probability. In 0.01 of all deliveries, these products are sold together. The lift value is 1.26.
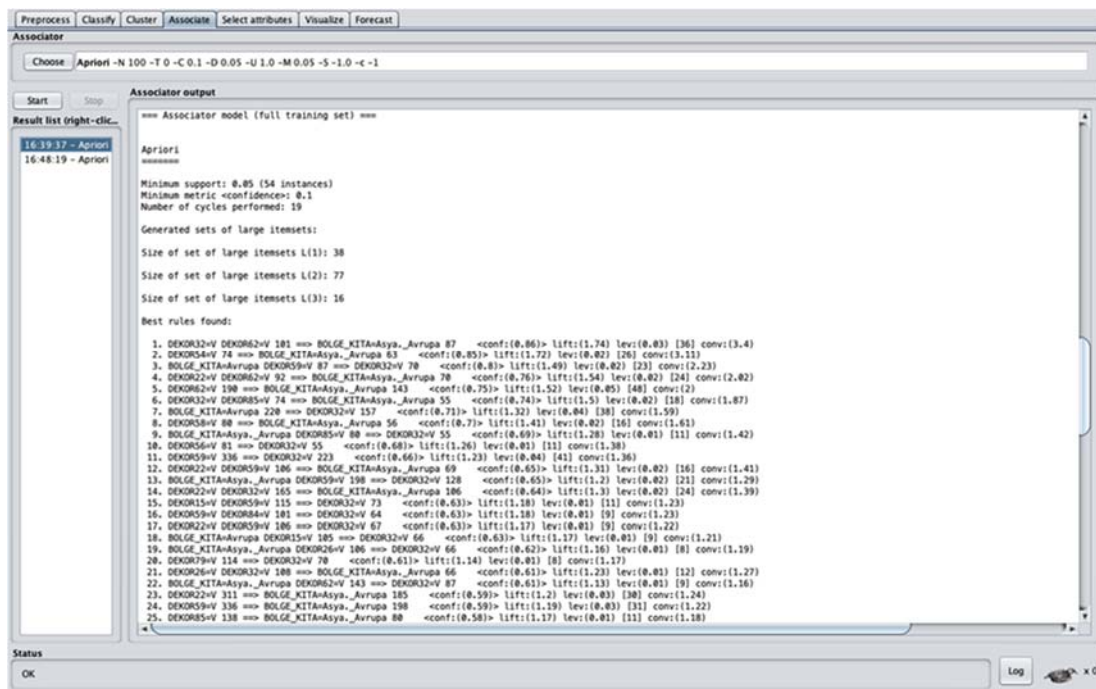

Figure 6. Apriori analysis results for dataset –2

When the rules formed in the association analysis are analyzed, it is observed that there is a trend effect. It is seen that the associations have changed in the data sets between 2009-2021 and 2019-2021. In addition, it is seen that the domestic and international rules are different. This is an indication that preferences and trends change on a country basis.

If the first 10 rules formed as a result of the analysis of the foreign sales data between 2009-2021, which is called data set - 1, are evaluated, it is quite possible to say that DECOR84 is a very popular

pattern abroad. The fact that all the rules are present and that the confidence intervals are so high confirms this opinion. It may be important to give priority to this pattern when proposing patterns to new companies at international fairs, color chart submissions, and to gain customers. Or, all foreign companies that are already customers of the company and purchase MCPB in preliminary designs within the framework of association rules and do not buy from DECOR84 are potential DECOR84 customers. Sending DECOR84 samples to these customers with the premise patterns or making promotions can greatly increase the sales

rates of DECOR84. In addition, companies that buy from DECOR84 pattern and not from predecessor decor papers can be researched and suggested these patterns.

When the first 10 rules formed as a result of the analysis of foreign sales data between the years 2019-2021, called data set - 2, are evaluated, it is seen that DECOR84 is not in the first 10 rules, unlike Data series - 1. Here it can be concluded that the trend has changed. The combinations of the patterns sold to the free zone are striking. Promotions and campaigns can be organized for customers in this region within the framework of these rules. Another remarkable point is the presence of DECOR32 in 6 different rules.

Also, although the association analysis is applied on the same patterns in two different data sets, the rules differ in the analyzes made on products that change according to current trends and tastes such as decor paper. This shows that not only the methods used, but also the periodic data sets are important in order to obtain the best results in the analysis.

5. CONCLUSION

Marketing activities are very important in a period where customer retention and acquisition of new customers require more effort than in the past. It is no longer enough to produce quality and low-cost products alone. Understanding customers and producing solutions to their needs is vital in order to gain a foothold in the market and compete with competitors.

Creating an accurate and effective marketing plan, as well as learning customer behaviors and demands, is a subject that companies strive for. Most of the companies have marketing units that work on these issues. It is the chief task of these units to determine the best marketing activities with the least cost and effort and to create an integrated marketing plan. All of the marketing activities such as surveys, market research, experienced marketing expert supplements, face-to-face customer interviews and advertisements are important for

companies to create a marketing plan, but they are also large cost items. Data return rate is more important than the number of people reached by the studies carried out in marketing activities. The return rate is a proof of how successful the marketing activity is for the company and that it brings customers. Increasing the rate of return is possible with the marketing work done to the right customer at the right time. Data mining is one of the ways to reach the right customer at the right time with the right campaign or product. Today, almost all companies use an ERP or CRM software. Analyzing and making sense of the accumulated data set thanks to the software used is indispensable for companies. Meaningful and correctly analyzed data sets allow companies to interpret the wishes, purchasing behaviors and trends of their current and potential customers. One of the data mining algorithms that reveals the purchasing behavior of customers and related purchases are the Apriori algorithm used in association analysis.

The aim of this study is to interpret the rules formed by the association analysis performed using the data sets prepared for analysis and to determine the purchasing tendencies and behaviors of the customers on the decor paper patterns. The orders received between the years 2009-2021 for the selected 88 decor papers were withdrawn from the company's ERP system, and a data set was prepared to be suitable for analysis and analyzed with the Apriori algorithm in the WEKA application. The results obtained will shed light on the company to know its customers and to create an efficient marketing strategy.

In the future, in addition to this study with foreign sales data, an association analysis can be made using domestic sales data and the resulting rules can be compared to investigate how customer behaviors differ.

6. REFERENCES

1. Shi, Z., Wang, G., 2018. Integration of Big-Data ERP and Business Analytics (BA). The Journal of High Technology Management Research, 29(2), 141-150.

762

Ç.Ü. Müh. Fak. Dergisi, 38(3), Eylül 2023

2. Han, J., Kamber, M., Pei, J., 2012. Data mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems, Elsevier, USA,740.

3. Angeline, D.M.D., 2013. Association Rule Generation for Student Performance Analysis Using Apriori Algorithm. The SIJ Transactions on Computer Science Engineering & its Applications (CSEA),1(1), 12-16.

4. Qiao, L., Peng, C., Guo, X., Wang, Y., 2017. Price Association Analysis of Agricultural Products based on Apriori Algorithm. Proceedings of Science. Information Science and Cloud Computing (ISCC 2017), China, 1-7.

5. Guo, Z., Chi, D., Wu, J., Zhang, W., 2014. A New Wind Speed Forecasting Strategy Based on The Chaotic Time Series Modelling Technique and The Apriori Algorithm, Energy Conversion and Management, 84, 140-151.

6. Zulfikar, W.B., Wahana, A., Uriawan, W., Lukman, N., 2016. Implementation of Association Rules with Apriori Algorithm for Increasing The Quality Of Promotion. In 2016 4th International Conference on Cyber and IT Service Management, 1-5, IEEE.

7. Xi, J., Zhao, Z., Li, W., Wang, Q., 2016. A Traffic Accident Causation Analysis Method based on AHP-Apriori. Procedia Engineering, 137, 680-687.

8. Sevri, M., Karacan, H., Akcayol, M. A., 2017. Crime Analysis based on Association Rules using Apriori Algorithm, International Journal of Information and Electronics Engineering, 7(3), 99-102.

9. Mirmozaffari, M., Alinezhad, A., Gilanpour, A., 2017. Data Mining Classification Algorithms for Heart Disease Prediction. Int'l Journal of Computing, Communications&Instrumentation Eng, 4(1), 11-15.

10. Cho, S.G., Kim, S.B., 2017. A Data-Driven Text Similarity Measure based on Classification Algorithms. International Journal of Industrial Engineering: Theory, Applications, and Practice, 24 (3), 328-339.

11. Beheshtian-Ardakani, A., Fathian, M., Gholamian, M., 2018. A Novel Model for Product Bundling and Direct Marketing in E-Commerce based on Market Segmentation. Decision Science Letters, 7(1), 39-54.

12. Zainol, Z., Wani, S., Nohuddin, P.N., Noormanshah, W.M., Marzukhi, S., 2018. Association Analysis of Cyberbullying on Social Media using Apriori Algorithm. International Journal of Engineering & Technology, 7, 72-75.

13. Kurnia, Y., Isharianto, Y., Giap, Y.C., Hermawan, A., 2019. Study of Application of Data Mining Market Basket Analysis for Knowing Sales Pattern (Association of Items) at the O! Fish Restaurant using Apriori Algorithm. In Journal of Physics: Conference Series, 1175 (1), 012047, IOP Publishing.

14. Hidayat, A.A., Rahman, A., Wangi, R.M., Abidin, R.J., Fuadi, R.S., Budiawan, W., 2019. Implementation and Comparison Analysis of Apriori and Fp-Growth Algorithm Performance to Determine Market Basket Analysis in Breiliant Shop. In Journal of Physics: Conference Series, 1402(7), 077031, IOP Publishing.

15. Simanjorang, R.M., Sijabat, P. 2020. Implementation of Apriori Algorithm in Determining the Level of Printing Needs. Infokum, 8(2), 43-48.

16. Dorling, L., Carvalho, S., Allen, J., Gonzalez-Neira, A., Luccarini, C., Wahlström, C., Rookus, M.A., 2021. Breast Cancer Risk Genes-Association Analysis in More than 113.000 Women. The New England Journal of Medicine, 384(5), 428-439.

17. Edastama, P., Bist, A.S., Prambudi, A., 2021. Implementation of Data Mining on Glasses Sales using The Apriori Algorithm. International Journal of Cyber and IT Service Management, 1(2), 159-172.

18. Pan, T., 2021. An Improved Apriori Algorithm for Association Mining Between Physical Fitness Indices of College Students. International Journal of Emerging Technologies in Learning (IJET), 16(9), 235-246.

19. Ünvan, Y.A., 2021. Market Basket Analysis with Association Rules. Communications in Statistics-Theory and Methods, 50(7), 1615-1628.

20. Zhao, X., Keikhosrokiani, P., 2022. Sales Prediction and Product Recommendation Model Through User Behavior Analytics. Computers, Materials, & Continua, 3855-3874.

21. Mirhashemi, S.H., Mirzaei, F., 2022. Using Combined Clustering Algorithms and Association Rules for Better Management of The Amount of Water Delivered to The Irrigation Network of Abyek Plain Iran, Neural Computing and Applications, 34(5), 3875-3883.

22. Aggarwal, C.C., 2015. Data Mining: The Textbook, Springer, New York, 1, 734.

23. Linoff, G.S., Berry, M.J., 2011. Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management. John Wiley & Sons, USA, 672.

24. Agrawal, R., Srikant, R., 1994. Fast Algorithms for Mining Association Rules. In Proc. 20th int. Conf. Very Large Data Bases (VLDB), 1215, 487-499.

25. Yuan, X., 2017. An Improved Apriori Algorithm for Mining Association Rules. In AIP Conference Proceedings, 1820(1), 080005.

26. Bodon, F., 2003. A Fast Apriori Implementation. In FIMI, 3, 63.

27. Koç, K.H., Dilik, T., Kurtoğlu, A., 2017. A Strategic Overview of the Turkish Forest Products Industry. IV. National Forestry Congress, 15, 16.

28. Yomsad, 2022. http://yomsad.org.tr/, Access date: 10.03.2022.