

## A STUDY ON THE USE OF DATA MINING IN THE PLANNING OF INVESTMENT FIELD

### Yatırım Alanlarının Planlamasında Veri Madenciliğinden Yararlanılması Üzerine Bir Çalışma

Ali Rıza İNCE\* & Mehmet Ali ALAN\*\*

#### Abstract

One of the most important strategic decisions taken by company managers at corporate level is in which areas to invest and how to manage and organize investments. Strategic management proposes portfolio analysis techniques in this regard. It is stated that portfolio analysis techniques, which have been criticized at many points, should be considered as initial techniques and should be supported by other techniques in practice. At this point, the question of whether or not data mining techniques will be used to identify new investment areas and allocate resources has constituted the main problematic of this study. In this study, it has been investigated whether there are association rules between investment areas thus, it is tried to reach the evaluations that may affect the strategic decisions in the process of determining investment areas, by using data belonging to various investor organizations. For this purpose, Association Rules Mining was conducted using data from 102 holding companies. As a result of the study, 35 rules were produced above the 50% confidence level. It is presented as a suggestion to the enterprises in which they can benefit from these rules in their investment planning.

#### Öz

Firma yöneticilerinin kurumsal düzeyde aldığı en önemli stratejik kararlardan biri, hangi alanlara yatırım yapılacak ve yatırımların nasıl yönetilip düzenleneceğidir. Stratejik yönetim, bu konuda portföy analiz tekniklerini önermektedir. Bir çok noktada eleştirilen portföy analiz tekniklerinin başlangıç teknikleri olarak değerlendirilmesi ve uygulamada diğer tekniklerle desteklenmesi gerektiği ifade edilmektedir. Bu noktada veri madenciliği tekniklerinin yeni yatırım alanlarının tespiti ve kaynakların tahsisi konusunda kullanılıp kullanılmayacağı sorusu bu çalışmanın temel sorunsalını oluşturmuştur. Bu çalışmada çeşitli yatırımcı kuruluşlara ait veriler kullanılarak, yatırım alanları arasında bireklik kurallarının olup olmadığı araştırılmış, böylelikle yatırımcı kuruluşlara yatırım alanları belirleme sürecinde, stratejik kararlarına etkileyebilecek değerlendirmelere ulaşımaya çalışılmıştır. Bu amaçla 102 holdinge ait veriler kullanılarak Bireklik Kuralları Madenciliği yapılmıştır. Yapılan çalışma sonucunda %50 güven seviyesinin üstünde 35 kural üretilebilmiştir. İşletmelerin Yatırım planlamalarında, bu kurallardan yararlanabilecekleri, işletmelere bir öneri olarak sunulmuştur.

**Anahtar Kelimeler:**  
Stratejik Yönetim,  
Yatırım Alanı,  
Planlama, Veri  
Madenciliği,  
Bireklik Kuralları.

**JEL Kodları:**  
M10, G11, C81

\* Associate Prof., Sivas Cumhuriyet University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, arince@cumhuriyet.edu.tr, ORCID: 0000-0003-4653-3091, Corresponding Author.

\*\* Prof., Sivas Cumhuriyet University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, alan@cumhuriyet.edu.tr, ORCID: 0000-0001-8562-547X

## **1. Introduction**

One of the most important strategic decisions taken by the managers at the institutional level of the companies which have several enterprises is which fields to invest and how to manage and organize the investments (Porter, 1987; Lynch and Smith, 2006: 5). The roles of the corporate center are searching for potentially high profitable businesses, purchasing, making them suitable for the organization and distributing strategic resources among these businesses (Thompson and Martin, 2010: 739). In other words; to define a collection of business investments that offers the most attractive opportunity for balancing cash flows and maximizing total returns in the long term (Roney, 2004: 215). At this point, corporate-level strategy in multi-enterprise companies requires senior management to adopt a portfolio management approach (Roney, 2004: 218).

Portfolio management techniques have been criticized by many authors (Slater and Wensley, 1981; Ansoff et al., 1982; Hambrick and Macmillan, 1982; Seeger, 1984; Hespelagh, 1989: 145; Aaker, 1992: 67; Slater and Zwirlein, 1992; Armstrong and Brodie, 1994;) and it is expressed that these techniques should be considered as initial techniques and should be supported by other techniques in practice (Day, 1977; Nippa et al., 2011). Identification of new investment fields of data mining and whether or not it will be used for allocating the resources have constituted the main research question of this study. Using data mining association relationship techniques, it can be determined whether the investment behavior of competitors creates a pattern.

Identifying and understanding these patterns can make a significant contribution to strategic management in which fields could be invested, how they can organize and manage their investments. It may be important for the companies investing in different fields to uncover associations between the fields invested. For this purpose, data mining techniques can be used to reveal the associations between the fields to be invested. In this study, it has been investigated whether there are association rules between investment fields by using data belong to various investor organizations thus, it is tried to reach the evaluations that may affect the strategic decisions of investor organizations in the process of determining investment fields.

The use of data mining applications in enterprises is increasing in recent years. Data mining is used in many fields such as quality control, error analysis, production systems and processes, maintenance, revenue growth and design of product and product family (Choudhary et al., 2009). This study, which aims to contribute to the identification and creation of the investment fields of the enterprises based on the applications of competitors, is an example of an application that was different and not previously made in terms of data mining. The study consists of four parts. Investment planning in the first part, in the second part, data mining, Association Rules technique and Apriori algorithm are discussed. The third part consists of the literature summary and the information about data set and method were given in the fourth part. Application related to the data of the investor companies were included in the fifth section.

## **2. Investment Planning**

While making investment decisions, companies benefit from analyzes such as Swot Analysis, E-V-R Congruence, Planning Gap Analysis, Porter's Industry Analysis and Competitive Advantage Frameworks, Scenario Modelling, Break-Even Analysis, Investment Appraisal Techniques Using Discounted, Cash Flows, Net Present Value, Internal Rate of Return, Payback and Cash-Flow Implications (Roney, 2004). However, when they need to make more

than one investment decision in related or unrelated fields, these analyzes are not sufficient to evaluate the impact of investments on the other investments and the whole company.

Strategic management research makes few proposals on methods used to effectively organize and manage multiple business portfolios that are vital for almost all medium-sized or large companies (Nippa et al., 2011). These proposals include portfolio analysis in which a large number of matrices are used such as BCG's Growth-share Matrix, Product/market Evolution Portfolio Matrix and Directional Policy matrix (DPM). Portfolio analysis can be generally useful in assessing the businesses or strategic business units of a company whose investments have diversified or both in terms of factors such as industry characteristics and market conditions. However, portfolio analysis may provide partial contribution on the subject of evaluation of current and future activity fields according to various criteria, more effective distribution of resources by seeing development potential in each field of activity, formulating business unit strategy and analysing portfolio balance (Grant and Jordan, 2015: 258).

Nevertheless, these analysis techniques have been criticized on the basis that multi-enterprise company management cannot make reliable decisions based on only a few variables (Wensley, 1981; Ansoff et al., 1982; Hambrick and Macmillan, 1982; Seeger, 1984), have difficulties in defining variables (Aaker, 1992: 29) and fail to address the issue of new business fields successfully (Hespelagh, 1989: 145). Portfolio analyzes, which aim to achieve high performance with lowest possible risk and creating strategic business segments that are necessary to achieve this (Kotler et al., 2010: 38), has been criticized at the point of disregarding interdependence and synergy issues (Thompson and Martin, 2010: 427). However, it has been evaluated as simplified methods that can only refer to ordinary strategic decisions (Slater and Zwirlein, 1992; Armstrong and Brodie, 1994). It has been recommended as a starting point that can be useful in strategic analysis (Day, 1977).

It is stated that corporate portfolio management, which is at the heart of the corporate strategy, includes the corporate strategic decision like introduction of new businesses, allocation of scarce resources and liquidation of value-destructive divisions, should not be limited to simple matrices or other tools used to manage the corporate portfolio and combination of portfolio techniques with other qualitative and quantitative analyzes is proposed (Nippa et al., 2011).

Data mining techniques that contribute to companies' determination of product portfolios (İnce and Alan, 2014) can also be helpful in taking strategic decisions at the institutional level, such as access to new business fields, allocation of scarce resources to different business units, and liquidation of non-generating units (Nippa et al., 2011). These techniques can be considered as inspiring useful diagnostic techniques just like portfolio matrices.

### **3. Data Mining, Association Rules and Apriori Algorithm**

Data mining is a combination of an interdisciplinary field and a set of disciplines, including database systems, statistics, machine learning, visualization, and information science (Han et al., 2006: 29). Data mining is used to provide preliminary information for decision support systems for the study area by discovering the information in databases (Fayyad et al., 1996). With data mining, it is possible to reveal the trends and behavior patterns necessary for decision support systems for the companies to make more effective decisions (İnan, 2003). Data mining, which makes it possible to discover information based on data analysis from different perspectives, has

the chance to be applied to a wide range of fields, including topics advertising, bioinformatics, database marketing, fraud detection, e-commerce, health, security, web, financial forecast (Jain et al., 2011). In addition, data mining is part of the overall process of discovering information in databases, as science and technology of discovery for uncovering previously unknown patterns. In today's computer-driven world, the vast amount of data in databases also includes patterns to be discovered. Accessibility and abundance of this information makes data mining very important and necessary (Rokach and Maimon, 2008: 1).

There are many techniques in data mining such as Association Rules, Clustering, Decision Trees, Discriminant Analysis, Artificial Neural Networks, Genetic Algorithms. These techniques are used to process information from various fields for the purpose of discovering and providing information which can guide a manager's decisions (Wu and Li, 2003).

The Association Rules technique has been developed in the field of computer science but it is mainly applied in fields such as market basket analysis (measuring the relationship between the products that a particular customer buys) and web click analysis (revealing the relationships between pages frequently clicked by a visitor for a website). Generally, the aim is to underline the part groups that often occur together in a group processes (Giudici and Figini, 2009: 90-91). In other words, Association Rules are the identification of relationships and associations in between special values of variables in large datasets. This technique allows analysts and researchers to reveal the hidden patterns in large data sets (Nisbet et al., 2009: 126). Revealing all possible patterns, easy to understand and usefulness is the strength of the technique. However, revealing all possible patterns is also a weakness of it. Because decision-makers will have to overcome a great deal of knowledge to evaluate all of these possibilities, which is difficult and time-consuming (Kantardzic, 2003: 169).

With the Association Rules technique, which is also called as Market Basket Analysis, evaluations are made for the analysis of consumption habits of more customers and it enables the identification of products or product groups that have a tendency to coexist in the purchasing process (Giudici and Figini, 2009: 175). Briefly, by the Association Rules, customers' consumption habits are revealed, and the customer is able to see other products that are likely to be bought while purchasing a product.

Algorithms such as Apriori, Fp-Growth, Tertius are used in the association rules technique. Apriori algorithm was used in this study. This algorithm finds rules based on validation values and uses first-order logical representations. These values include various options such as class index, class validation threshold, validation value, frequency threshold, loss values and data, noise threshold, number patterns, repeat patterns, ROC analysis and output values. Apriori is an inductive logic programming algorithm, seeking the results of the validation evaluation function with the highest values. It explores the simplest different validation measures, which are weighted with relative accuracy. A validation criterion indicates part of the unexpected and expected counter examples of a rule (Arora et al., 2013). Two values that have the expected and observed probability are calculated in the algorithm. Apriori extracts first-order rules and is used in Association Rules mining missions with other programs (Nahar et al., 2013). General structure of Apriori Algorithm is as follows (Webb, 2003):

1.  $L_1 = \{\text{frequent one-item sets}\}$
2. for  $k = 2$ ;  $L_{k-1} \neq \emptyset$ ;  $k++$  do begin
3.  $C_k = \{\{x_1, x_2, \dots, x_{k-2}, x_{k-1}, x_k\} \mid \{x_1, x_2, \dots, x_{k-2}, x_{k-1}\} \in L_{k-1} \wedge$

```

 $\{x_1, x_2, \dots, x_{k-2}, x_k\} \in L_{k-1}$ 
4. for all_transactions  $t \in D$  do begin
5. for all candidates  $c \in C_k \wedge c \subseteq t$  do
6.  $c.count++;$ 
7. end
8.  $L_k = \{c \in C_k \mid c.count \geq minsup\}$ 
9. End
10. return  $U_k L_k$ ;
    
```

The two statistics initially used in association rules are the support and confidence level. These are numerical values and some numerical terms must be defined to identify these.

“Let  $D$  be the database of transactions and  $N$  be the number of transactions in  $D$ . Each transaction  $D_i$  is an item set. Let  $Support(X)$  be the proportion of transactions that contain item set  $X$ :

$$Support(X) = |\{I \mid I \in D \wedge I \supseteq X\}| / N \quad (1)$$

where  $I$  is an item set and  $|\cdot|$  denotes the cardinality of a set. The support of an association rule is the proportion of transactions that contain both the antecedent and the consequent. The confidence of an association rule is the proportion of transactions containing the antecedent that also contain the consequent. For an association  $A \rightarrow C$ ,

$$Support(A \rightarrow C) = Support(A \cup C) \quad (2)$$

$$Confidence(A \rightarrow C) = Support(A \cup C) / Support(A) \quad (3)$$

If support value is sufficiently high (and the transactions represent a random sample from the same data distribution as future transactions), then confidence is a reasonable estimate of the probability that any given future transaction that contains the antecedent will also contain the consequent” (Webb, 2003: 27-28).

Another value in the rules of the union is the lift value. The lift value is calculated as follows:

$$Lift(A \rightarrow C) = Confidence(A \rightarrow C) / Support(C) \quad (4)$$

If the result of the operation is less than 1, it means that seeing  $A$  has a negative correlation with seeing  $C$ . If it is greater than 1, it means that seeing  $A$  has a positive correlation with seeing  $C$ , which means that seeing one is associated with seeing the other. If the lift value is 1, it means that the two sides are independent of each other.

There are several popular algorithms used in the production of association rules. Among these, Apriori and FP-Growth are the most known algorithms. In this study, Apriori algorithm was used to determine the association rules. The Apriori algorithm is one of the most preferred and advanced sequential algorithms. This algorithm creates candidate items using only item sets that have been identified as large in a previous scan without using the transactions in database. The Apriori algorithm is based on the assumption that any subset of a large set of items will also be large. Thus, a set of items consisting of  $k$  items can be obtained by combining large sets of items with  $k-1$  items and deleting the ones with non-large subsets. At the end of this combining and deleting transaction, fewer candidate item sets will be comprised (Silahtaroğlu, 2016: 142).

#### **4. Literature Summary**

There are many studies on different data sets in the relevant literature. Among these, Stiloul et al. (2001) applied the Apriori algorithm to a database containing records of diabetic patients and issued rules of association from hidden patterns. They have shown that the extracted results may be a good value for diagnostic procedures, especially in the case of large data volumes. Kumar and Rukmani (2010) have produced association rules using Apriori and FP-Growth algorithms from web log files. Umarani and Punithavalli (2011) analyzed the real-life data such as retail sales data and market basket data with different association rules mining algorithms. Ilayaraja and Meyyappan (2013) showed the rules of frequency of diseases and association between diseases by using apriori algorithm with the data of 29 recurrent diseases in people living in different time zones and regions. Bansal and Bhambhu (2013) produced association rules using the Apriori algorithm from a set of data of crime related to women. In the same study, the performances of apriori and Predictive Apriori Algorithms were compared, and it has been shown that Apriori is better and faster than the Predictive Apriori Algorithm. Angeline (2013), produced association rules by analyzing relationship between different attributes and student achievements with Apriori Algorithm with the student data. İnce and Alan (2014) determined that there is a rule of association between different funding tools when rival enterprises acquiring funds.

Guo et al. (2014) found that the rules of association between wind speed and other meteorological factors could be effectively explored by the apriori algorithm in their study on the estimation of wind velocity by using wind speed and other meteorological data. In their study, Sutisnawati and Reski (2019) analyzed transaction data in restaurants and revealed the relationships between products purchased by consumers with the Apriori algorithm. Hong et al. (2020) tried to discover hidden patterns and potential associations in risk factors using crash data of freight trucks from 19,038 accidents on Korean highways from 2008 to 2017. For this purpose, they produced 90,951 association rules using the Apriori algorithm. Simanjorang (2020) found that the most ordered items were Cake Boxes and Brochures, by looking at the products that provide minimum support and minimum trust using the apriori algorithm, using the data of the most ordered printed materials. Adero et al. (2020) mined association rules using district-based crime data in Kenya, and stated that these discovered rules could be used to determine the crime risk factor of districts. Nan and Chen (2020), using the physical fitness test data of university and college students, discovered the potential relationship between the Apriori algorithm and the indicators in the physical fitness tests, and offered suggestions for improving performance in this way. Qisman et al. (2021) searched for association rules among the products sold using data from the computer store, and presented the 22 most preferred rules. Zeng et al. (2021), using data obtained from 2,621 USGS indicators in the United States between 1981 and 2016, determined the relationships between annual stream flow trends and precipitation trends with the Apriori algorithm. As can be seen, the rules of data mining and, more specifically, association rules are the techniques used to reveal predictable or unpredictable patterns in every field of science.

#### **5. Data Set and Method**

In the study, 102 company which operates under a holding title in Turkey were evaluated. In the study, all holdings in Turkey were included by using the full count method. The fields where these companies operate have been categorized one by one according to the guideline of classifying professions used by Istanbul Chamber of Commerce (ICC). The reason for using ICC's

classification guideline is that it contains fewer titles than other guidelines. Fewer titles have made it easier to reach association rules. Ethics committee approval and/or legal/special permission were not required in this study, and research and publication ethics were complied with. Association Rules Mining was conducted by using this data. In this context, data warehouse was prepared by using 8.262 data about 81 different investment fields of 102 holding companies by using Excel macros and was subjected to analyze.

The data warehouse is the sum of the subject-oriented, time-variant and non-updateable data used in management decision support. According to Anahory and Murray (1997), a data warehouse is meta, fact, dimensional and aggregational data and is the process manager that allows people to make informed decisions and provides appropriate information. A partial screenshot of the original version of the dataset is as follows:

	A	B	C	D	E	F	G	H	I	J	
1	Holdingler	PERAKENDE TİCARET	TOPTAN VE DIŞ TİCARET	BÜTÜNLERİ, OYUNCAK VE HEDİYELİK EŞYA	KUYUMCULUK	BİLGİ TEKNOLOJİLERİ	KAĞIT VE KİRTASİYE	KOZMETİK	İLAÇ VE TİBBİ CİHAZ	TOPTAN GIDA VE TEMİZLİK ÜRÜNLERİ	CAM VE ÇİHAZ
2	Holding 1										
3	Holding 2										
4	Holding 3										
5	Holding 4										
6	Holding 5										
7	Holding 6						1				
8	Holding 7										
9	Holding 8							1	1		
10	Holding 9					1					
11	Holding 10	1					1			1	
12	Holding 11										
13	Holding 12									1	
14	Holding 13									1	
15	Holding 14						1	1	1		1
16	Holding 15										1
17	Holding 16										1
18	Holding 17						1				
19	Holding 18										
20	Holding 19										
21	Holding 20										
22	Holding 21										
23	Holding 22						1				
24	Holding 23										

**Figure 1. Screenshot of the Original Version of Dataset**

During the preparation of the data warehouse in this study, the value “1” was assigned if an investment has been made in any investment field, “0” value was assigned if there is no investment in this field. In the study, all investment fields are defined as variable and have taken values of {1,0}. A part of the screenshot of the data warehouse obtained after editing the dataset is as follows:

**Figure 2. Screenshot of the Data Warehouse**

## 6. Application

In the study, version 1.4.5 of the Tanagra software was used. Tanagra is an open source software. This software supports the algorithms of many classification, clustering and Association Rules. By using the current data, the Apriori algorithm supported by the Tanagra program produced 35 rules over 50% confidence level. The results produced by Apriori Algorithm are presented in Table 1. The second column of the table shows the first condition and the third column shows the second condition of the rules. The fourth column in the table shows the support value, and the fifth column shows the confidence level. In the following table, 35 rules with 50% or more confidence level are presented.

**Table 1. Rules Produced By Apriori Algorithm**

N	Antecedent	Consequent	Support (%)	Confidence (%)	Lift
1	"Underwear_and_Accessories=True"	"Non-Ferrous_Metals=True" - "Carpet_and_Floor_Coverings=True"	12.871	68.421	2.8793
2	"Non-Ferrous_Metals=True" - "Carpet_and_Floor_Coverings=True"	"Underwear_and_Accessories=True"	12.871	54.167	2.8793
3	"Underwear_and_Accessories=True"	"Fuel-Oil=True"	10.891	57.895	2.4364
4	"Iron_and_Steel=True"	"Animal_Food_Products=True"	10.891	64.706	2.4204
5	"Non-Ferrous_Metals=True" - "Underwear_and_Accessories=True"	"Carpet_and_Floor_Coverings=True"	12.871	92.857	2.4047
6	"Fuel-Oil=True"	"Non-Ferrous_Metals=True" - "Carpet_and_Floor_Coverings=True"	11.881	50.000	2.1041
7	"Non-Ferrous_Metals=True" - "Carpet_and_Floor_Coverings=True"	"Fuel-Oil=True"	11.881	50.000	2.1041
8	"Underwear_and_Accessories=True"	"Carpet_and_Floor_Coverings=True"	14.851	78.947	2.0445
9	"Carpet_and_Floor_Coverings=True" - "Underwear_and_Accessories=True"	"Non-Ferrous_Metals=True"	12.871	86.667	2.0356
10	"Non-Ferrous_Metals=True" - "Animal_Food_Products=True"	"Carpet_and_Floor_Coverings=True"	10.891	78.571	2.0348
11	"Carpet_and_Floor_Coverings=True" - "Property_Consultants=True"	"Non-Ferrous_Metals=True"	9.901	83.333	1.9573
12	"Non-Ferrous_Metals=True" - "Fuel-Oil=True"	"Carpet_and_Floor_Coverings=True"	11.881	75.000	1.9423
13	"Carpet_and_Floor_Coverings=True" - "Fuel-Oil=True"	"Non-Ferrous_Metals=True"	11.881	80.000	1.8790
14	"Animal_Food_Products=True"	"Property_Consultants=True"	13.861	51.852	1.8703
15	"Property_Consultants=True"	"Animal_Food_Products=True"	13.861	50.000	1.8703
16	"Non-Ferrous_Metals=True" - "Property_Consultants=True"	"Carpet_and_Floor_Coverings=True"	9.901	71.429	1.8498
17	"Underwear_and_Accessories=True"	"Non-Ferrous_Metals=True"	13.861	73.684	1.7307
18	"Carpet_and_Floor_Coverings=True" - "Animal_Food_Products=True"	"Non-Ferrous_Metals=True"	10.891	73.333	1.7224
19	"Jewelry=True"	"Traffic_Consultancy=True"	9.901	62.500	1.6611
20	"Fuel-Oil=True"	"Carpet_and_Floor_Coverings=True"	14.851	62.500	1.6185
21	"Financial_Institutions=True"	"Traffic_Consultancy=True"	13.861	60.870	1.6178
22	"Hotelshome_Textile=True"	"Traffic_Consultancy=True"	12.871	59.091	1.5705

**Table 1. Continue**

N	Antecedent	Consequent	Support (%)	Confidence (%)	Lift
23	"Fuel-Oil=True"	"Non-Ferrous_Metals=True"	15.842	66.667	1.5658
24	"Textile_Finishing=True"	"Non-Ferrous_Metals=True"	12.871	65.000	1.5267
25	"Hotels=True"	"Non-Ferrous_Metals=True"	9.901	62.500	1.4680
26	"Carpet_and_Floor_Coverings=True"	"Non-Ferrous_Metals=True"	23.762	61.538	1.4454
27	"Non-Ferrous_Metals=True"	"Carpet_and_Floor_Coverings=True"	23.762	55.814	1.4454
28	"Animal_Food_Products=True"	"Carpet_and_Floor_Coverings=True"	14.851	55.556	1.4387
29	"Textile_Finishing=True"	"Carpet_and_Floor_Coverings=True"	10.891	55.000	1.4243
30	"Property_Consultants=True"	"Traffic_Consultancy=True"	14.851	53.571	1.4238
31	"HotelsHome_Textile=True"	"Carpet_and_Floor_Coverings=True"	11.881	54.545	1.4125
32	"Restoration_and_Isolation=True"	"Traffic_Consultancy=True"	11.881	50.000	1.3289
33	"Animal_Food_Products=True"	"Non-Ferrous_Metals=True"	13.861	51.852	1.2179
34	"Property_Consultants=True"	"Non-Ferrous_Metals=True"	13.861	50.000	1.1744
35	"Restoration_and_Isolation=True"	"Non-Ferrous_Metals=True"	11.881	50.000	1.1744

As can be seen from the table above, according to the first rule produced; the enterprises invest in the "Underwear\_and\_Accessories" field also invest in the "Non-Ferrous\_Metals" or "Carpet\_and\_Floor\_Coverings" fields with a confidence level of 68.421%. In other words, in 12.871% of the analyzed investment transactions, enterprises also invested in "Non-Ferrous\_Metals" or "Carpet\_and\_Floor\_Coverings" fields along with "Underwear\_and\_Accessories" field. 68.421% of the enterprises that invested in field "Underwear\_and\_Accessories" invested in fields "Non-Ferrous\_Metals" or "Carpet\_and\_Floor\_Coverings". The fact that the Lift value in the rule is greater than 1 indicates that investing in field "Underwear\_and\_Accessories" has a positive correlation with investing in fields "Non-Ferrous\_Metals" or "Carpet\_and\_Floor\_Coverings".

According to the second rule; the enterprises investing in the "Non-Ferrous\_Metals" or "Carpet\_and\_Floor\_Coverings" fields invest in the "Underwear\_and\_Accessories" field with a confidence level of 54.116%. In other words, in 12.871% of the analyzed investment transactions, enterprises also invested in "Underwear\_and\_Accessories" field along with "Non-Ferrous\_Metals" or "Carpet\_and\_Floor\_Coverings" fieldS. 68.421% of the enterprises that invested in fieldS "Non-Ferrous\_Metals" or "Carpet\_and\_Floor\_Coverings" invested in field "Underwear\_and\_Accessories". The fact that the Lift value in the rule is greater than 1 indicates that investing in fields "Non-Ferrous\_Metals" or "Carpet\_and\_Floor\_Coverings" has a positive correlation with investing in field "Underwear\_and\_Accessories".

According to the third rule; the enterprises investing in the "Underwear\_and\_Accessories" field invest in the "Fuel-Oil" field with a confidence level of 57,895%. In other words, in 12.871% of the analyzed investment transactions, enterprises also invested in "Fuel-Oil" field along with "Underwear\_and\_Accessories" field. 68.421% of the enterprises that invested in field "Underwear\_and\_Accessories" invested in field "Fuel-Oil". The fact that the Lift value in the rule is greater than 1 indicates that investing in field "Underwear\_and\_Accessories" has a positive correlation with investing in field "Fuel-Oil". For the remaining 32 rules, similar comments can be made.

## 7. Conclusion and Comments

One of the most important strategic decisions taken by the managers of the enterprises is deciding the fields in which the enterprises will invest. With this study, it was aimed to contribute to the process of defining and creating the investment fields of the enterprise by a strategic approach based on investment fields made by competitors in decision-making processes for business managers. In order to achieve the purpose of the study, the Association Rules technique of data mining was used.

The investment fields of the holdings operating in different fields have been examined and it was tried to determine the rules and patterns between the investment behaviors of these holdings by the Association Rules technique. In this way, it has been tried to contribute to the enterprises in the phase of defining and creating the fields to invest, and limitation of several fields and to keep them to an optimum level.

In this study, the data related to the investment fields of the holdings were compiled from the websites and analyzed with the Association Rules technique. In this context, 8.262 data of 81

fields where 102 holdings invested were analyzed by Apriori algorithm using version 1.4.5 of Tanagra software.

As a result of the analysis, 35 Association Rules was determined above the 50% confidence level in the choice of investment fields. According to the first rule produced; the enterprises investing in the "Underwear\_and\_Accessories" field also invest in the "Non-Ferrous\_Metals" or "Carpet\_And\_Floor\_Coverings" fields with a confidence level of 68,421%. 34 rules which are similar to this rule have been identified.

According to these results, it is possible to establish a suggestion system in which any investor company will be reminded of possible fields of investment in defining and creating the investment field according to the rules produced.

**Declaration of Research and Publication Ethics**

This study which does not require ethics committee approval and/or legal/specific permission complies with the research and publication ethics.

**Researchers' Contribution Rate Statement**

The authors declare that they have contributed equally to the article.

**Declaration of Researcher's Conflict of Interest**

There are no potential conflicts of interest in this study.

## References

- Aaker, D. (1992). *Strategic marketing management*. New Work: Wiley.
- Adero, E., Okeyo, G. and Mwangi, W. (2020, May). *Using apriori algorithm technique to analyze crime patterns for Kenyan national crime data: A county perspective*. In M. Cunningham and P. Cunningham (Eds.), *IST-Africa 2020* (pp. 140-148). Paper presented at The 2020 IST-Africa Conference (IST-Africa). Retrieved from <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9144029>
- Anahory, S. and Murray, D. (1997). *Data warehousing in the real world: A practical guide for building decision support systems*. Harlow, UK: Addison-Wesley
- 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. Retrieved from <http://www.thesij.com/>
- Ansoff, H.A., Kirsch, W. and Roventa, P. (1982). Dispersed positioning in portfolio analysis. *Industrial Marketing Management*, 11(4), 237-252. [https://doi.org/10.1016/0019-8501\(82\)90013-X](https://doi.org/10.1016/0019-8501(82)90013-X)
- Armstrong, J.S. and Brodie, R.J. (1994). Effects of portfolio planning methods on decision making: Experimental results. *International Journal of Research in Marketing*, 11(1), 73-84. [https://doi.org/10.1016/0167-8116\(94\)90035-3](https://doi.org/10.1016/0167-8116(94)90035-3)
- Arora, J., Bhalla, N. and Rao, S. (2013). A review on association rule mining algorithms. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(5), 1246-1251. Retrieved from <https://www.ijircce.com/>
- Bansal, D. and Bhambaru, L. (2013). Execution of apriori algorithm of data mining directed towards tumultuous crimes concerning women. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(9), 54-62. Retrieved from <http://www.ijarcs.info/index.php/Ijarcs>
- Choudhary, A.K., Harding, J.A. and Tiwari, M.K. (2009). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5), 501-521. doi:10.1007/s10845-008-0145-x
- Day, G.S. (1977). Diagnosing the product portfolio. *Journal of Marketing*, 41(2), 29-38. <https://doi.org/10.1177/002224297704100213>
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37-54. <https://doi.org/10.1609/aimag.v17i3.1230>
- Giudici, P. and Figini, S. (2009). *Applied data mining for business and industry* (2nd Ed.). West Sussex: Wiley Publication.
- Grant, R.M. and Jordan, J.J. (2015). *Foundations of strategy*. UK: John Wiley & Sons.
- Guo, Z., Chi, D., Wu, J. and 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(1), 140-151. <https://doi.org/10.1016/j.enconman.2014.04.028>
- Hambrick, D. and Macmillan, I. (1982) The product portfolio and Man's best friend. *California Management Review*, 21(1), 84-95. <https://doi.org/10.2307/41164995>
- Han, J., Kamber, M. and Pei, J. (2006). *Data mining, Southeast Asia edition: Concepts and techniques*. USA: Elsevier.
- Haspeslagh, P. (1989). Portfolio planning: Uses and limits. In D. Asch and C. Bowman (Eds.), *Readings in strategic management* (pp. 144-161). London: Palgrave.
- Hong, J., Tamakloe, R. and Park, D. (2020). Discovering insightful rules among truck crash characteristics using apriori algorithm. *Journal of Advanced Transportation*, 2020, 4323816. <https://doi.org/10.1155/2020/4323816>

- Ilayaraja, M. and Meyyappan, T. (2013). Mining medical data to identify frequent diseases using apriori algorithm. In Department of Computer Science (Ed.), *Pattern recognition, informatics and mobile engineering* (pp. 194-199). India: IEEE
- İnan, O. (2003). *Veri madenciliği* (Yayınlanmamış yüksek lisans tezi). Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya.
- İnce, A.R. and Alan, M.A. (2014). Ürün portföy planlamasında veri madenciliğinden yararlanması üzerine bir çalışma. *EUL Journal of Social Sciences*, 2, 64-77. Retrieved from <http://euljss.eul.edu.tr/>
- Jain, Y., Kumar, V., Kumar Y. and Geetika S.P. (2011). An efficient association rule hiding algorithm for privacy preserving data mining. *International Journal on Computer Science and Engineering*, 3(7), 2792-2798. Retrieved from <http://www.enggjournals.com/ijcse/>
- Kantardzic, M. (2003). *Data mining: Concepts, models, methods, and algorithms*. New Jersey: John Wiley & Sons.
- Kotler, P., Berger, R. and Bickhoff, N. (2010). *The quintessence of strategic management. What you really need to know to survive in business*. Berlin: Springer.
- Kumar, B.S. and Rukmani, K.V. (2010). Implementation of web usage mining using apriori and fp-growth algorithms. *International Journal of Advanced Networking and Applications*, 1(6), 400-404. Retrieved from <https://www.ijana.in/>
- Lynch, R.L. and Smith, J.R. (2006). *Corporate strategy*. Harlow: Prentice Hall.
- Nahar, J., Imam, T., Tickle, K.S. and Chen, Y.P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, 40(4), 1086-1093. <https://doi.org/10.1016/j.eswa.2012.08.028>
- Nan, S. and Chen, M. (2020). *An apriori-algorithm-based analysis method on physical fitness test data for college students* (EasyChair Working Paper No. 4522). Retrieved from <https://yahootechpulse.easychair.org/publications/preprint/vM6g>
- Nippa, M., Pidun, U. and Rubner, H. (2011). Corporate portfolio management: Appraising four decades of academic research. *Academy of Management Perspectives*, 25(4), 50-66. <https://doi.org/10.5465/amp.2010.0164>
- Nisbet, R., Elder, J. and Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Burlington: Elsevier.
- Porter, M. (1987). *From competitive advantage to corporate strategy*. New York: The Free Press.
- Qisman, M., Rosadi, R. and Abdullah, A.S. (2021). Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (Case study of Mizan computer retail stores). *Journal of Physics: Conference Series*, 1722(1). doi:10.1088/1742-6596/1722/1/012020
- Rokach, L. and Maimon, O. (2008). *Data mining with decision trees: Theory and applications*. New Jersey: World Scientific.
- Roney, C.W. (2004). *Strategic management methodology: Generally accepted principles for practitioners*. Connectitut: Greenwood Publishing Group.
- Seeger, J.A. (1984). Research note and communication. Reversing the images of BCG's growth/share matrix. *Strategic Management Journal*, 5(1), 93-97. <https://doi.org/10.1002/smj.4250050107>
- Silahtaroğlu, G. (2016). *Veri madenciliği kavram ve algoritmaları* (3. bs.). İstanbul: Papatya Bilim.
- Simanjorang, R.M. (2020). Implementation of apriori algorithm in determining the level of printing needs. *Infokum*, 8(2), 43-48. Retrieved from <http://infor.seaninstiute.org/>
- Slater, S.F. and Zwirlein, T.J. (1992). Shareholder value and investment strategy using the general portfolio model. *Journal of Management*, 18(4), 717-732. <https://doi.org/10.1177/014920639201800407>
- Stiloul, S., Bamidis, P.D., Maglaveras, N. and Pappas, C. (2001). Mining association rules from clinical databases: An intelligent diagnostic process in healthcare. In V. Patel et al. (Eds.), *Studies in health technology and informatics* (pp. 1399-1403). Amsterdam: IOS Press.

- Sutisnawati, Y. and Reski, M. (2019). *Looking for transaction data pattern using apriori algorithm with association rule method*. Paper presented at the IOP Conference Series: Materials Science and Engineering. doi:10.1088/1757-899X/662/2/022078
- Thompson, J.L. and Martin, F. (2010). *Strategic management: Awareness and change*. UK: Cengage Learning Emea.
- Umarani, V. and Punithavalli, M. (2011). An empirical analysis over the four different methods of progressive sampling-based association rule mining. *European Journal of Scientific Research*, 66(4), 620-630. Retrieved from <https://www.europeanjournalofscientificresearch.com/>
- Webb, G.I. (2003). Association rules. In N. Ye (Ed.), *The handbook of data mining* (pp. 27-28). New Jersey: Lawrence Erlbaum Associates.
- Wensley, R. (1981). Strategic marketing: Betas, boxes, or basics. *The Journal of Marketing*, 45(3), 173-182. <https://doi.org/10.1177/002224298104500314>
- Wu, T. and Li, X. (2003). Data storage and management. In N. Ye (Ed.), *The Handbook of data mining* (pp. 393-407). New Jersey: Lawrence Erlbaum Associates.
- Zeng, X., Schnier, S. and Cai, X. (2021). A data-driven analysis of frequent patterns and variable importance for streamflow trend attribution. *Advances in Water Resources*, 147, 103799. <https://doi.org/10.1016/j.advwatres.2020.103799>