



Received: August 08, 2019  
Accepted: December 22, 2019  
Published Online: December 31, 2019

AJ ID: 2018.07.02.MIS.01  
DOI: 10.17093/alphanumeric.630866  
**Research Article**

## A Study On Profiling Students via Data Mining

Mehmet Ali Alan, Ph.D. \* 

Assoc. Prof., Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Cumhuriyet University, Sivas, Turkey, [alan@cumhuriyet.edu.tr](mailto:alan@cumhuriyet.edu.tr)

Mustafa Temiz 

Res. Assist., Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Cumhuriyet University, Sivas, Turkey, [temizmustafa@cumhuriyet.edu.tr](mailto:temizmustafa@cumhuriyet.edu.tr)

\* Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi 58140, Kampüs, Sivas, Türkiye

### ABSTRACT

Data mining is a significant method which is utilized in order to reveal the hidden patterns and connections within big data. The method is used at various fields such as financial transactions, banking, education, health sector, logistics and security. Even though analysis towards the consumption habits of the customers is carried out via association rules mining more often, which is one of the basic methods of data mining, the method is also utilized in order to profile patients and students. As well as the customization of a customer is of high significance, so is distinguishing and customizing a student. Within this study, students were tried to be profiled via data mining of the student data of a high school. A set of qualities, that can directly affect the performance of students such as health conditions, financial resources, life standards and education level of the families, were taken into consideration. For that purpose, upon the analysis of data of 443 students in the database, a data warehouse was established. The Apriori algorithm, which is one of the popular algorithms of association rules mining, is utilized for the data analysis. Apriori algorithm was able to produce 72 rules which are accurate above 90%. It is thought that the produced rules can be of help in profiling the students, and they can contribute to work of school management, teachers, parents and students.

### Keywords:

Data Mining, Association Rules, Data Warehouse, Student Profile



## 1. Introduction

Today, a significant amount of data is stored within the databases of social media and institutions. Besides getting and protecting the data, evaluation of it is also of high importance. One of the media in which the data is kept is school databases. Analysis of the data within these databases and the knowledge that can be derived out of it can be of great help for both school management and teachers.

Families can be reluctant in sharing some information with the school management because of some certain reasons. This behavior may result in lack of information about the student and thus some negative effect on the success level. Student profiling studies can contribute to the connection between the school and the family.

Data mining methods can discover helpful information that can be utilized within forming evaluation in order to be of help in terms of establishing a pedagogical base when the educators design or modify an educational approach or environment (Romero and Ventura, 2007).

The data mining, which is generally defined as the discovery of information embedded into databases, is mainly known by its significant role in revealing secret data out of large amounts of data. Application of data mining is utilized in e-trade, bio-informatics, and educational studies which are recently known as educational data mining (Mohamad and Tasir, 2013).

Within this study which aims to profile students, via the association rules mining method, student profiling out of the data of 443 high school students was tried to be carried out.

Study consists of four sections. In the first section data mining, association rules and Apriori algorithm were explained, in the second section literature abstract was presented and in the third section data set and methods were explained. In the final section the analysis of the data of 443 high school students is given.

## 2. Data Mining, Association Rules and Apriori Algorithm

Data mining is the process of discovering and extracting the information from covered patterns, databases and data warehouses. A variety of algorithms and tools, which can be utilized for this purpose, exist (Parack et. al., 2012).

Data mining is a part of the main process which aims to discover data within databases as a research science for the discovery of previously unknown patterns. In our computer based world, such databases cover enormous amounts of information. Availability and abundance of the knowledge makes the data mining significant and necessary (Rokach and Maimon, 2008).

A variety of data mining algorithms, such as association rules, piling, decision trees, discriminant analysis, neural network and genetic algorithms, exist. These algorithms could be used with the purpose of information processing from a series of fields in order to discover information which can alter the decision making processes of executives. Information is the data which is connected to current day and the past. Information maintains the ground for future trends which are based on original data

and essential information which can be derived from original data. Frankly, knowledge and information are related via data (Wu and Li, 2003: 393 – 407).

Association rules are one of the frequently exercised data mining methods which target revealing association conditions which occur within large sets of data (Birant et. al., 2010: 215 – 221; Agrawal et. al., 1993: 207 – 216).

Association rules are developed in the field of computer sciences and frequently used in order to measure the correlation among the products a certain customer purchases , and to measure the connection among the websites which are surfed by internet users, within important applications like market basket. Mainly, the purpose is highlighting the element groups which typically appear with a set of processes. The data, on which the association rules are applied, is the data in which the processes are stored in a database form. For each action (a line in the database), the database includes the list of actualized elements.

Each individual can be seen within a set of data more than once. With the market basket analysis, transaction means a single visit to the supermarket in which the shopping list is saved; with the web click analysis, transaction means a web login in which all of the visited web page lists are recorded. Lines cover different numbers of elements and this difference is significant to the data matrix. As an alternative to the previously mentioned methods, the database can be converted into double data matrix via processes like lines, transactions and columns (Giudici and Figini, 2009: 90 – 91).

In the association rules mining, apriori algorithm turned into a standard approach. They were introduced by Agrawal and Srikant (1994) for the first time. Algorithm initiates with a pile of data which covers transactions, and targets building a set of products which are over the threshold of one single customer at least. In the algorithmic process of apriori, an element of group X with a length k is defined as frequent when the all X subgroups of k length are frequent. This kind of evaluation decreases the search area significantly and results in the discovery of some certain rules. Confidence, mainly stands for the accuracy of the rule and is utilized to line the rules up within Apriori (Nahar et. al., 2013).

The main steps of apriori algorithm are as follows (Webb, 2003):

Apriori Algorithm;

```

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.  $c.\text{count}++;$ 
7. end
8.  $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
9. end
10. return  $U_k L_k;$ 

```

The two initial statistics to association rules are support value and confidence level. They are numerical values and some numerical terminology must be defined prior to the definition of these values. Given that  $D$  is the database of the transactions and  $N$  is the number of transactions in  $D$ . Each  $D_i$  transaction is a group of products. *Support* ( $X$ ) is the ratio of transactions which covers the product group  $X$ :

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

$I$  is a group of elements and  $|I|$  shows the element number of the group.

The support value of this association rule is the ratio of former and latter transactions to number of total transactions. Confidence value, on the other hand, is the accuracy ratio of determining the latter via utilizing the former. For the  $A \Rightarrow C$  association, confidence and support values are as follows (Webb, 2003):

$$Support(A \Rightarrow C) = Support(A \cup C)$$

$$Confidence(A \Rightarrow C) = Support(A \cup C) / Support(A)$$

If the support value is high enough (if the transactions represent a random sample of same data distribution with future transactions) the confidence level is an acceptable estimation of any future transaction which covers the first side of the rule, and possibility of the rule to cover the second side as well (Webb, 2003).

Besides support and confidence measurements within the association rules, another calculation method is the "lift" value. Lift value is calculated as follows:

$$Lift(A \Rightarrow C) = \frac{Confidence(A \Rightarrow C)}{Support(C)}$$

If the result is smaller than 1, getting the result A has negative correlation over getting result C; if the result is bigger than 1, getting the result A has positive correlation over getting result C and it means that getting the results are significantly related. If the lift value is 1, then both sides are independent from each other (Taş, 2018: 37-38).

### 3. Literature Abstract

A variety of studies exist which were carried out covering similar data sets. These are;

With the purpose of understanding, estimating and preventing academic failure among university students, Bresfelean et. al. (2008), presented scientific studies based on data mining methods which cover past scholastic situations and survey data. In order to achieve this goal, authors carried out analysis, which are based on classification and grouping methods, to profile the students for their failure/success in taking the exams (Bresfelean et. al., 2008).

In their study, Parack et. al., (2012) discussed the application of student profiling and grouping in education and data mining. Via utilizing Apriori algorithm, which is one of the most popular association rules mining, they profiled students. With the purpose of grouping the students, they did their groupings by K-means algorithm which assigns a set of observations to sub-groups and found out that the preferred

algorithms, which can be applied within education systems, are efficient ways of profiling students (Parack et. al., 2012).

At a Turkish university, Aydemir (2019), tried to estimate the grades of students who took the Foreign Language 102 class via utilizing data mining methods. The data set of this study consists of the data of 3974 students. In terms of the grades, 12 qualities, which can differ in the student profile, such as department, curriculum type, faculty, day or night school, students' grade in placement, placement order and average grades to the class in previous year, were included. As the classifier to the study which utilized the tenfold cross verification method, the performances of Bagging, Neural Webs, M5Rules, DecisionStump, DecisionTable and M5P algorithms were checked against each other. By these algorithms, the Bagging method was found to be the most successful one with 0.80 correlation coefficient and 1.22 absolute mistake ratios (Aydemir, 2019).

Within his/her study, Angeline (2013) used the Apriori algorithm in order to classify the students according to their school success levels and to build group of rules related to these classes. In that study, the factors that affect the success level of the students were tried to be determined and solutions which can be of help in terms of increasing the success level were proposed. Including gender, department, education level of the parents, self confidence level of the student and financial status, 15 self-quality info was used. Relation rules between self-qualities and classes were evaluated in three levels as good, bad and moderate. By the Apriori algorithm, which is association rule mining, 127 rules were created (Angeline, 2013).

Gara and Pado (2015), tried to determine the departments that can be preferred by students in accordance with their characteristics and used profile info of the students for self qualities. With this study, the purpose was being of help to high school students and administrative staff in terms of making successful decisions in further education. Student profiles were analyzed by the Apriori algorithm which is one of the association rules mining methods. 6 rules were developed and the most successful results were 81,36%, 61,29% and 1,03 for confidence, support and lift respectively (Gara and Pado, 2015).

Huang et. al., (2018), utilized the Apriori algorithm in order to reveal association rules for the self-qualities of 164 B.A degree students according to their grades and discipline competition rewards. By analysis, it was found out that the students with higher grades of C# development, towards object, internet web designing, data structure (C#) and basic programming classes, have a higher susceptibility to win the discipline competitions (Huang, 2018).

Within their study, which includes knowledge of the students about the topic, the ability of learning with helping units, the teachers' motivation towards self and students, teachers' communicative skills, teacher's control over the class, timing, order and knowledge beyond the curriculum, Singh et. al., (2011), utilized grouping and association rules. They have developed 14 rules and with this study which uses student feedback, they have run a success analysis for a whole faculty (Singh, 2011).

Using the student profiles, Zawayda (2013) carried out a study that targets to be of help to students in making decisions for students in getting registered to M.A degree education. Thanks to the abundance of data in M.A degree database, he/she utilized

the Apriori algorithm in order to discover required data which can be helpful for students and universities. The data used in the study consists of the answers of 207 students from 15 different departments including finance, banking and tourism to surveys. A set of 8 self-qualities were evaluated including employment type, religion, lifestyle, gender, financial status, marital status, profession and advices from friends. Via the apriori algorithm, 19 rules were generated (Zawayda, 2013).

#### 4. Dataset and Method

In this study, the data set was gathered from a high school in Sivas province of Turkey. Association rules mining was applied to this data set. In this context, the data of 443 students was appropriately converted by Excel macro and the data warehouse was established.

Data warehouse is the sum of subject based, time related and non-updateable data which is used as a support unit for administrative decision making processes. In terms of structure, data warehouse is a data of commodity, real, dimensional and grouped characteristics, and it is a process director in the sense of helping people in taking knowledgeable decisions and providing appropriate information (Bose, et. al., 2009: 190). Within our study, fully trained/educated data set was utilized.

#### 5. Application

The 3.7.2. version of the Weka software was used in the study. Weka is an open source software. The software supports algorithms which are connected to many classifying, grouping and association rules.

In the study, students were tried to be profiled by utilizing Apriori algorithm. For this purpose, the self-quality data of high school students which was given in Table 1 was used to analyze the associations.

Line	Variable	1	0
1	Father_Deceased	Yes	No
2	Mother_Deceased	Yes	No
3	Father_Step	Yes	No
4	Mother_Step	Yes	No
5	None of the parents work	Yes	No
6	Parents are divorced	Yes	No
7	Father_Illiterate	Yes	No
8	Mother_Illiterate	Yes	No
9	Father is often away because of his job	Yes	No
10	Father is in prison	Yes	No
11	Number of siblings	Few	Many
12	Number of sibligs attending to university	Yes	No
13	Disabled sibling	Yes	No
14	Chronic diseases	Yes	No
15	There are permanent guests at home	Yes	No
16	Amount of funds	Sufficient	Insufficient
17	Working after school	Yes	No
18	Financial support is required	Yes	No
19	Accepted to university	Yes	No

**Table 1.** Variables and Assigned Values

By the analysis 72 rules with confidence level of 90% or over could be generated. Results were presented in Table 2.

According to the results given in Table 2, the left side of the mark “=>” shows the first condition of the rule and the right hand side shows the second condition. The “Conf” mark stands for the confidence level. And the lift value lies at the end of the rule. In the following table, the rules which are generated via application of study data set.

1. Mother_Illiterate=1 Working after school=1 14 => Financial support is required=1 14 <conf:(1)> lift:(4.87)
2. None of the parents work=1 Mother_Illiterate=1 Number of sibligs attending to university=1 11 => Financial support is required=1 11 <conf:(1)> lift:(4.87)
3. None of the parents work=1 Mother_Illiterate=1 Working after school=1 11 => Financial support is required=1 11 <conf:(1)> lift:(4.87)
4. None of the parents work=1 Number of sibligs attending to university=1 Working after school=1 10 => Financial support is required=1 10 <conf:(1)> lift:(4.87)
5. None of the parents work=1 Father_Illiterate=1 9 => Financial support is required=1 9 <conf:(1)> lift:(4.87)
6. Mother_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 9 => Financial support is required=1 9 <conf:(1)> lift:(4.87)
7. Father_Illiterate=1 Working after school=1 8 => Financial support is required=1 8 <conf:(1)> lift:(4.87)
8. None of the parents work=1 Father_Illiterate=1 Mother_Illiterate=1 8 => Financial support is required=1 8 <conf:(1)> lift:(4.87)
9. None of the parents work=1 Mother_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 8 => Financial support is required=1 8 <conf:(1)> lift:(4.87)
10. Father_Illiterate=1 Mother_Illiterate=1 Working after school=1 7 => Financial support is required=1 7 <conf:(1)> lift:(4.87)
11. Father_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 7 => Financial support is required=1 7 <conf:(1)> lift:(4.87)
12. Father_Deceased=1 Working after school=1 6 => Financial support is required=1 6 <conf:(1)> lift:(4.87)
13. None of the parents work=1 Father_Illiterate=1 Working after school=1 6 => Mother_Illiterate=1 6 <conf:(1)> lift:(11.66)
14. None of the parents work=1 Father_Illiterate=1 Working after school=1 6 => Financial support is required=1 6 <conf:(1)> lift:(4.87)
15. Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 6 => Working after school=1 6 <conf:(1)> lift:(11.08)
16. Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 6 => Financial support is required=1 6 <conf:(1)> lift:(4.87)
17. Number of sibligs attending to university=1 There are permanent guests at home=1 Working after school=1 6 => Financial support is required=1 6 <conf:(1)> lift:(4.87)
18. None of the parents work=1 Father_Illiterate=1 Working after school=1 Financial support is required=1 6 => Mother_Illiterate=1 6 <conf:(1)> lift:(11.66)
19. None of the parents work=1 Father_Illiterate=1 Mother_Illiterate=1 Working after school=1 6 => Financial support is required=1 6 <conf:(1)> lift:(4.87)
20. None of the parents work=1 Father_Illiterate=1 Working after school=1 6 => Mother_Illiterate=1 Financial support is required=1 6 <conf:(1)> lift:(16.41)
21. Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 Financial support is required=1 6 => Working after school=1 6 <conf:(1)> lift:(11.08)
22. Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 6 => Financial support is required=1 6 <conf:(1)> lift:(4.87)
23. Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 6 => Working after school=1 Financial support is required=1 6 <conf:(1)> lift:(14.29)
24. Mother_Step=1 There are permanent guests at home=1 5 => Mother_Deceased=1 5 <conf:(1)> lift:(20.14)
25. Mother_Deceased=1 Number of sibligs attending to university=1 Amount of funds=1 5 => There are permanent guests at home=1 5 <conf:(1)> lift:(6.33)
26. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 5 => Mother_Illiterate=1 5 <conf:(1)> lift:(11.66)
27. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 5 => Working after school=1 5 <conf:(1)> lift:(11.08)
28. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 5 => Financial support is required=1 5 <conf:(1)> lift:(4.87)
29. Mother_Illiterate=1 There are permanent guests at home=1 Working after school=1 5 => Financial support is required=1 5 <conf:(1)> lift:(4.87)
30. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 5 => Mother_Illiterate=1 5 <conf:(1)> lift:(11.66)
31. None of the parents work=1 Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 5 => Working after school=1 5 <conf:(1)> lift:(11.08)
32. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 5 => Mother_Illiterate=1 Working after school=1 5 <conf:(1)> lift:(31.64)
33. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 Financial support is required=1 5 => Mother_Illiterate=1 5 <conf:(1)> lift:(11.66)
34. None of the parents work=1 Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 5 => Financial support is required=1 5 <conf:(1)> lift:(4.87)
35. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 5 => Mother_Illiterate=1 Financial support is required=1 5 <conf:(1)> lift:(16.41)
36. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 Financial support is required=1 5 => Working after school=1 5 <conf:(1)> lift:(11.08)
37. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 5 => Financial support is required=1 5 <conf:(1)> lift:(4.87)
38. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 5 => Working after school=1 Financial support is required=1 5 <conf:(1)> lift:(14.29)
39. None of the parents work=1 Number of sibligs attending to university=1 Working after school=1 Accepted to university=1 5 => Financial support is required=1 5 <conf:(1)> lift:(4.87)
40. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 Financial support is required=1 5 => Mother_Illiterate=1 5 <conf:(1)> lift:(11.66)
41. None of the parents work=1 Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 Financial support is required=1 5 => Working after school=1 5 <conf:(1)> lift:(11.08)
42. None of the parents work=1 Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 5 => Financial support is required=1 5 <conf:(1)> lift:(4.87)
43. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 Financial support is required=1 5 => Mother_Illiterate=1 Working after school=1 5 <conf:(1)> lift:(31.64)
44. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 Working after school=1 5 => Mother_Illiterate=1 Financial support is required=1 5 <conf:(1)> lift:(16.41)
45. None of the parents work=1 Father_Illiterate=1 Mother_Illiterate=1 Number of sibligs attending to university=1 5 => Working after school=1 Financial support is required=1 5 <conf:(1)> lift:(14.29)
46. None of the parents work=1 Father_Illiterate=1 Number of sibligs attending to university=1 5 => Mother_Illiterate=1 Working after school=1 Financial support is required=1 5 <conf:(1)> lift:(31.64)
47. Father_Deceased=1 Chronic diseases=1 4 => There are permanent guests at home=1 4 <conf:(1)> lift:(6.33)
48. Parents are divorced=1 Chronic diseases=1 4 => Number of sibligs=0 4 <conf:(1)> lift:(2.24)
49. Parents are divorced=1 Disabled sibling=1 4 => Accepted to university=1 4 <conf:(1)> lift:(2.67)
50. Disabled sibling=1 Financial support is required=1 4 => Father_Illiterate=1 4 <conf:(1)> lift:(22.15)
51. Father_Illiterate=1 Disabled sibling=1 4 => Financial support is required=1 4 <conf:(1)> lift:(4.87)
52. Amount of funds=1 Financial support is required=1 4 => Number of sibligs=0 4 <conf:(1)> lift:(2.24)
53. Chronic diseases=1 Working after school=1 4 => Number of sibligs attending to university=1 4 <conf:(1)> lift:(2.18)

54. Father_Deceased=1 There are permanent guests at home=1 Working after school=1 4 ==> Financial support is required=1 4 <conf:(1)> lift:(4.87)
55. Mother_Deceased=1 Number of siblings=0 Amount of funds=1 4 ==> There are permanent guests at home=1 4 <conf:(1)> lift:(6.33)
56. Mother_Deceased=1 There are permanent guests at home=1 Financial support is required=1 4 ==> Working after school=1 4 <conf:(1)> lift:(11.08)
57. None of the parents work=1 Mother_Illiterate=1 There are permanent guests at home=1 4 ==> Financial support is required=1 4 <conf:(1)> lift:(4.87)
58. Mother_Illiterate=1 Working after school=1 Accepted to university=1 4 ==> None of the parents work=1 4 <conf:(1)> lift:(11.36)
59. None of the parents work=1 Number of sibilgs attending to university=1 Chronic diseases=1 4 ==> Financial support is required=1 4 <conf:(1)> lift:(4.87)
60. Father_Illiterate=1 Number of sibilgs attending to university=1 Accepted to university=1 4 ==> Financial support is required=1 4 <conf:(1)> lift:(4.87)
61. Mother_Illiterate=1 Working after school=1 Accepted to university=1 4 ==> Financial support is required=1 4 <conf:(1)> lift:(4.87)
62. Father is often away because of his job=1 Number of siblings=0 Number of sibilgs attending to university=1 4 ==> Amount of funds=1 4 <conf:(1)> lift:(2.18)
63. None of the parents work=1 Mother_Illiterate=1 Number of sibilgs attending to university=1 Accepted to university=1 4 ==> Financial support is required=1 4 <conf:(1)> lift:(4.87)
64. Mother_Illiterate=1 Working after school=1 Financial support is required=1 Accepted to university=1 4 ==> None of the parents work=1 4 <conf:(1)> lift:(11.36)
65. None of the parents work=1 Mother_Illiterate=1 Working after school=1 Accepted to university=1 4 ==> Financial support is required=1 4 <conf:(1)> lift:(4.87)
66. Mother_Illiterate=1 Working after school=1 Accepted to university=1 4 ==> None of the parents work=1 Financial support is required=1 4 <conf:(1)> lift:(15.28)
67. There are permanent guests at home=1 Working after school=1 Financial support is required=1 Accepted to university=1 4 ==> None of the parents work=1 4 <conf:(1)> lift:(11.36)
68. Number of sibilgs attending to university=1 Working after school=1 18 ==> Financial support is required=1 17 <conf:(0.94)> lift:(4.6)
69. None of the parents work=1 Mother_Illiterate=1 17 ==> Financial support is required=1 16 <conf:(0.94)> lift:(4.58)
70. Mother_Deceased=1 Amount of funds=1 12 ==> There are permanent guests at home=1 11 <conf:(0.92)> lift:(5.8)
71. Father_Illiterate=1 Mother_Illiterate=1 10 ==> Financial support is required=1 9 <conf:(0.9)> lift:(4.38)
72. Father_Illiterate=1 Number of sibilgs attending to university=1 10 ==> Financial support is required=1 9 <conf:(0.9)> lift:(4.38)

**Table 2.** The Rules Generated via Apriori Algorithm

According to the first rule in table 2.14 out of 14 students whose mothers are illiterate need financial support. The confidence level of this rule is 1 and the lift value is 4.87.

According to the second rule, 11 out of 11 students whose both parents are unemployed, mothers are illiterate, and have siblings who attend to a university, need financial support. The confidence level to this rule is 1 and lift value is 4.87.

According to the third rule, 11 out of 11 students whose both parents are unemployed, mothers are illiterate, and works after school need financial support. The confidence level of this rule is 1 and the lift value is 4.87.

According to rule 4, 10 out of 10 students whose parents are unemployed, have sibling(s) that attend to a university, and works after school need financial support. The confidence level of this rule is 1 and the lift value is 4.87.

According to rule 5.9 out of 9 students whose parents are unemployed and fathers are illiterate need financial support. The confidence level of this rule is 1 and the lift value is 4.87.

Such interpretations can be done over remaining 67 rules.

## 6. Results

This study has the purpose of being help to school administrative, teachers and parents in terms of understanding the students better. In order to achieve the goals of the study, the association rules method of the data mining was utilized.

Via the association rules mining method over the data gathered from school databases, the connections between self-qualities of students were analyzed and students were tried to be profiled according to the self-qualities. In accordance with the outcomes, the info that aims defining the factors that may affect success levels of students, was determined.

In the study, the data set of a high school was analyzed via association rules mining method. In this context, the data of 443 students was analyzed via 3.7.2. version of Weka software which utilizes the Apriori algorithm.



By the analysis, 72 rules with confidence level of 90% or over could be generated. According to the first rule, 14 out of 14 students whose mothers are illiterate, and works after school need financial support. It was also found out that the confidence level of this rule is 1 and the lift value is 4,87. Similarly, other 71 rules were generated.

Based on the results, it seems obvious that data mining is crucial in terms of decision making for educational institutions. Analyzing the high amounts of data manually in order to define the pattern is painfully difficult. Instead of following such a demanding method, it is possible to profile the students via data mining methods, and within this study, a group of students, who completed their high school education, were profiled. Taking the generated rules into consideration is of high significance for school administration, teachers and parents in terms of understanding the students better and contributing to their success level.

## References

- Agrawal, R., Imielinski, T. & Swami, A. (1993). "Mining Association Rules between Sets of Items in Large Databases". *Acm sigmod record* 22 (1993) 207-216.
- 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 (2013) 12-16.
- Aydemir, E. (2019). Geçme Notlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi". *European Journal of Science and Technology* 15 (2019) 70-76. doi: <https://doi.org/10.31590/ejosat.518899>.
- Birant, D., Kut, A., Ventura, M., Altınok, H., Altınok, B., Altınok, E., & İhlamur, M. (2010). İş Zekâsı Çözümleri için Çok Boyutlu Birliktelik Kuralları Analizi". *Akademik Bilişim* 10 (2010) 256.
- Bose, I., Chun, L. A., Yue, L. V. W., Ines, L. H. W. & Helen, W. O. L. (2009). "Business Data Warehouse: The Case of Wal-Mart". *Data Mining Applications for Empowering Knowledge Societies*. Ed. Rahman H. Information Science Reference. (2009) 189-198. Bangladesh.
- Bresfelean, V. P., Bresfelean, M., Ghisoiu, N., & Comes, C. A. (2008). "Determining students' academic failure profile founded on data mining methods". *ITI 2008-30th International Conference on Information Technology Interfaces* (2008) 317-322. doi: 10.1109/ITI.2008.4588429
- Gara, G. P. P., & Padoa, F. R. F. (2015). "Mining Association Rules on Students Profiles and Personality Types". *Proceedings of the International Multiconference of Engineers and Computer Scientists 1* (2015) 307-312.
- Giudici, P. & Figini S. (2008). "Applied Data Mining For Business and Industry", A John Wiley and Sons, Ltd., Publication. 2008 90-91.
- Huang, X., Xu, Y., Zhang, S., & Zhang, W. (2018). "Association rule mining for selecting proper students to take part in proper discipline competition: a case study of Zhejiang University of Finance and Economics". *International Journal of Emerging Technologies in Learning (IJET)* 13 (2018) 100-113.
- Mohamad, S. K., & Tasir, Z. (2013). "Educational data mining: A review". *Procedia-Social and Behavioral Sciences* 97 (2013) 320-324.
- Nahar J., Imam T., Tickle K. S., Chen Y. P. (2013). "Association Rule Mining to Detect Factors Which Contribute To Heart Disease in Males and Females". *Expert Systems with Applications* 40 (2013) 1086-1093. doi:<https://doi.org/10.1016/j.eswa.2012.08.028>.
- Parack, S., Zahid, Z., & Merchant, F. (2012). "Application of data mining in educational databases for predicting academic trends and patterns". *2012 IEEE International Conference on Technology Enhanced Education (ICTEE)* (2012) 1-4. doi: 10.1109/ICTEE.2012.6208617.
- Rokach, Lior and Maimon, Oded (2008), *Data Mining with Decision Trees*, World Scientific, New Jersey
- Romero, C., & Ventura, S. (2007). "Educational data mining: A survey from 1995 to 2005". *Expert systems with applications* 33 (2007) 135-146.

- Singh, C., Gopal, A., & Mishra, S. (2011). "Extraction and analysis of faculty performance of management discipline from student feedback using clustering and association rule mining techniques". 2011 3rd International Conference on Electronics Computer Technology 4 (2011) 94-96). Doi: 10.1109/ICECTECH.2011.5941864
- Taş, Y. (2018). Birliktelik Kuralları Madenciliği ve Bir Uygulama, Master's Dissertation, Sivas Cumhuriyet University, Sivas 2018.
- Webb, G. I. (2003). "Association Rules". Ed. Ye N. The Handbook Of Data Mining. (2003) 27-28. New Jersey.
- Wu, T. & Li, X. (2003). "Data Storage and Management". Ed. Ye N. The Handbook of Data Mining. (2003) 393-407. New Jersey.
- Zawayda, Y. I. A. (2013). "Mining postgraduate students' data using apriori algorithm". Doctoral dissertation. Universiti Utara, Malaysia