ISSN 2349-6193
Impact Factor: 2.805

IJESMR

International Journal OF Engineering Sciences &Management Research

IMPROVING THE PERFORMANCE OF THE INSTITUTIONBY USING APRIORI ALGORITHM

Ahmet Çinar[1] & Muhammed Yildirim[*2]
[1&*2]Computer Engineering Department / Firat University, Turkey

Keywords: *data mining, association rules, apriori algorithm.*

ABSTRACT
Public institutions are units that are set up for public public services. In order to meet the needs of the society and to continue the social life on a regular basis, public services need to be carried out without disruption. Increasing the performance of public institutions will make it easier for the state and for the citizens to do their jobs easily. In this study, it is aimed to increase the performances of institutions by using apriori algorithm which is one of the algorithms for extracting association rules in data mining and to make the work of the citizens faster and easier. Moreover, in this study, the subjects that institutions should work together are observed. It is aimed to accelerate the operation of the institutions and to increase the performances of the personnel by detecting these issues

INTRODUCTION
With today's developing technology, the size of the data held in databases have been increasing day and day. This increment makes processing on the data complicated. For this reason, the extraction of previously unknown relationships using large data sets has an important place [1].Data mining is a process that has been developed to make important data meaningful within these data sets.Data mining has become one of the most popular and up-to-date technologies in recent years. Because of the increasingly cheaper computer systems and the increased power of computers, computers are able to store larger amounts of data. The fact that more than one process carried out due to the development of technology is recorded in electronic environment ensures that these recorded data can be safely stored and accessed easily when requested and that these processes can be obtained at a lower cost every passing day. For this reason, it is of utmost importance to be able to use methods that can handle these datapool[2].A large amount of data is held in databases of institutions or companies. However, institutions or companies have difficulties in extracting meaningful and useful information from these data sets. Data mining is the process of obtaining applicable and meaningful information that can be useful within all of these datapool. Data mining is the whole of data analysis techniques. It is a tool to obtain the information necessary to solve the current problem using the available data, to make important decisions or make forecasts for the future, and at the same time to reveal hidden patterns and rules in databases [3]. The databases of institutions are expanding with the developing technology. It is important to extract applicable and meaningful information that can be useful in all these datapool. Using existing data, it is an important task to solve the existing problem and to solve these problems[4].

METHOD
There are many methods and algorithms developed in the data mining process. In this study, the apriori algorithm, which is one of the data mining association rule extraction algorithms, is used to improve the performance of institutions or employees. The application is developed in Visual Studio (C # language) using Apriori algorithm. It has been tried to determine how to improve the performance of institutions or employees together with the deductions obtained in practice.

Rules of Association
Agrawal, Imielinski and Swami first addressed the problem of analysis of association rules in 1993[5]. It is the name given to the data mining methods that resolve the co-occurrence of events. The rules of association are used to describe co-occurring associations. It is the job of finding cohesion relationships among different data within large data sets. The rules of association provide for the summation and summarization of frequently recurring relations. Since the stored data is increasing day by day, it is necessary to determine the relations between the data in the database and each other. Various arrangements are made thanks to the obtained association rules.

ISSN 2349-6193
Impact Factor: 2.805

IJESMR
International Journal OF Engineering Sciences &Management Research

There are 2 main processes while forming the rules of association. In the first process, there are frequently repeated items. Each of these items has to be reprinted with a minimum number of predefined support. The second is to form association rules from frequently repeated items. When these rules are created, they must provide minimum support and minimum confidence [6]

Apriori Algorithm

The Apriori algorithm was developed by Agrawall and Srikant in 1994. The apriori algorithm has a wide range of applications, from education to banking, from engineering to telecommunications to telecommunication.The Apriori algorithm takes the word "prior" because it takes the information it uses from the previous step [7]. The basic rationale of the Apriori algorithm is that it has an iterative structure. It is used in the discovery process of frequent repetitive element clusters in databases. According to the Apriori algorithm, if the k-element cluster (element cluster with k elements) meets the minimum support criterion, the subset of this cluster also provides the minimum support value [8]. The rules of association used in the data mining process often include items that are frequently repeated in the database [9]. These items, which are then repeated frequently, produce association rules [5]. In the rules of association, the association between items is calculated taking into account support and trust values. The Support value indicates how much the relationship between the items in the database. Products A and B are different from each other; The support value for product A is the ratio of product A in all products [10].

Support (A) = Number of A products / Total number of shopping.
The support value for products A and B is the probability that A and B co-exist together at all exchanges.
Support (A, B) = Number of (A, B) products / Total number of products.
Confidence is the probability that product B will be associated with product A.
Confidence (A, B) = number of products containing products (A, B) / number of products containing product A.
At the same time,
Trust (A => B) = Support (A, B) / Support (A) [11]

The reliability of the association rules obtained is directly proportional to the support and trust values. Each rule is expressed with a confidence and support value (A => B [support value = 2%, confidence value = 60%]). It means that 2% of the support value of the association rule is sold together with 2% of all the products analyzed. A confidence level of 60% indicates that 60% of customers who buy product A also buy product B at the same Exchange [12].

APLICATION AND RESULTS

In this study, the subjects that the institutions should work together are observed. It is aimed to accelerate the operation of the institutions and to increase the performances of the personnel by detecting these issues. If the documents that come with this study are related to which institution, they will be guided by 12 different subject groups. We can show these subject groups as in Table 1.

Table 1: Table of paper subjects

1. Request	4. Job Request	7. Appointment Request	10. Officer Complaint
2. Thanks	5. Help Request	8. Complaint	11. Suggestion
3. Thanks to the institution	6. Staff Request	9. Institution Complaint	12. Opinion

Incoming documents need to receive the document number from the document receipt. We can show it as in Table 2.

Table2: Document Record Table

Document number 1	Document number 3	Document number 5	Document number 7
Document number 2	Document number 4	Document number 6	Document number 8

In the later stage, the stage in which the document goes to which stage. A part of the institutional table is symbolically as shown in Table 3.

Table3: Institution Table

A institution	C institution	E institution	G institution
B institution	D institution	F institution	H institution

ISSN 2349-6193
Impact Factor: 2.805

IJESMR
International Journal OF Engineering Sciences &Management Research

We can summarize the general operation as in Table 4.

Table4: Documents and Institutional Relationship

Document Number	Documents sending institutions	The subject of the paper
Document number 1	A institution, B institution	Request
Document number 2	C institution, B institution	Business request
Document number 3	E institution	Help
Document number 4	H institution, K institution	Thanks

- That the subject of document number 1 is request and this document is sent to institution A and institution B,
- That the subject of document number 2 is a business request and that this document has been sent to institution C and institution B,
- That the subject of document number 3 is help and this document is sent to institution E
- It is understood that the document number 4 is thank and this document has been sent to the institution H and the institution K.

In the application using the software for extraction of association is implemented in C# . In the database, document table, institution table and category table are created and associated separately.
All records are available on the tables and can be added or deleted by the user, as well as the document number, institutions and their relationships later in the application interface.

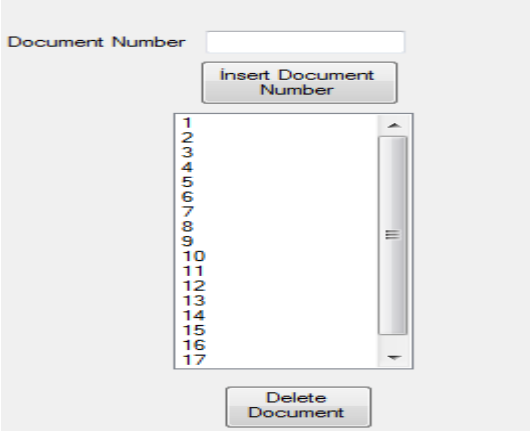

Figure1: Adding, deleting and viewing documents

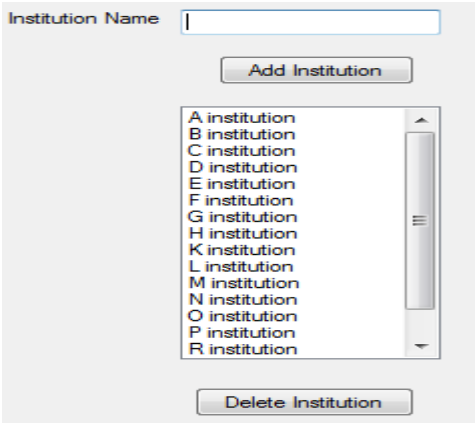

Figure 2: Adding and deleting institutions

ISSN 2349-6193
Impact Factor: 2.805

IJESMR
International Journal OF Engineering Sciences &Management Research

The user who adds the document number and / or institution can also enter the institutions sending documents. The corresponding interface is as in Figure 3.

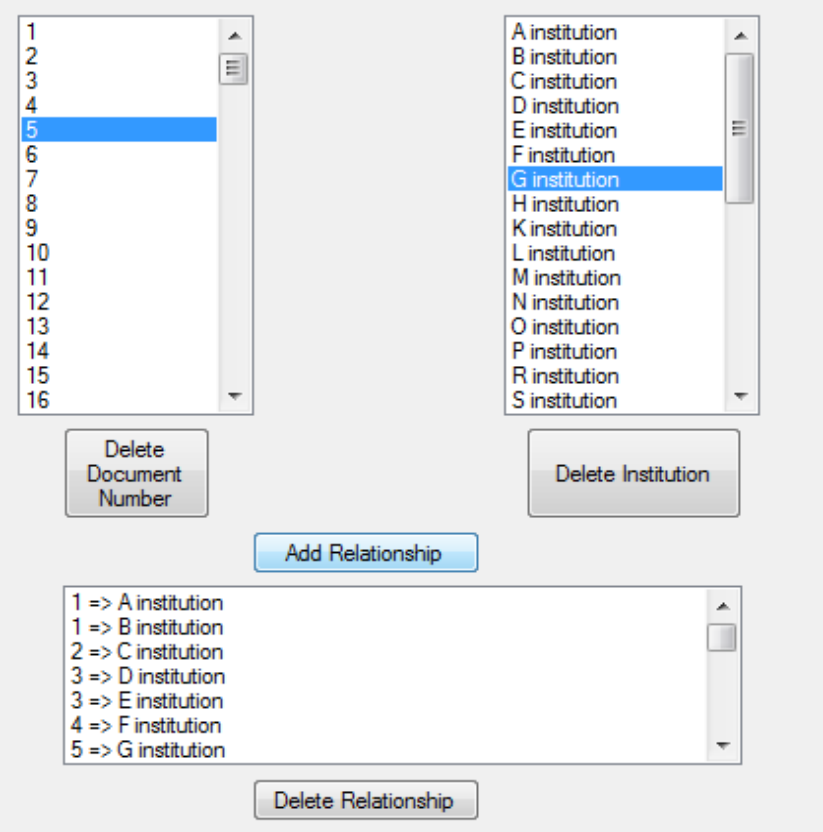

Figure 3: Associating the institution with the document

We understand Figure 1.3 goes to institution A and institution B, document 2 goes to institution C, document 3 goes to institution D and institution E, document 4 goes to institution F and document 5 goes to institution G. If the user asks for the document number and institution of the table and adds the relation, then he can add new relations.

Relations between institutions can be depicted as in Figure 4.

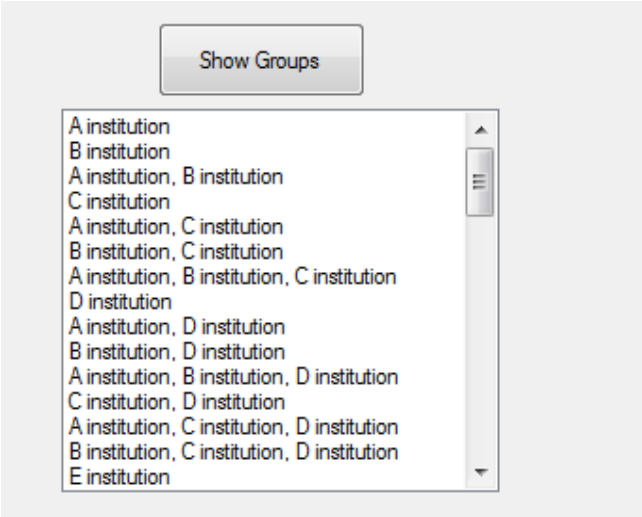

Figure4: Institutions list of documents

ISSN 2349-6193
Impact Factor: 2.805

IJESMR
International Journal OF Engineering Sciences &Management Research

The final stage is the stage in which the rules of association are drawn. This step is the same as in Table 5.

Table5:List of association rules

Institution Name	Number of Associations	The Subject of the Paper	Confidence Value(%22)	Support Value(%1)
A institution	25	Request		%8,77
B institution	10	Request		%3,5
C institution	70	Request		%24,5
D institution	109	Request		%38,2
E institution	65	Request		%22,8
F institution	5	Request		%1,75
G institution	8	Request		%2,8
H institution	3	Request		%1,05
J institution	30	Request		%10,52
K institution	15	Request		%5,26
L institution	5	Request		%1,75
A institution => B institution	6	Request	%24	%2,1
C institution =>D institution	39	Request	%55,7	%13,6
C institution =>E institution	30	Request	%42,8	%10,52
E institution =>D institution	52	Request	%80	%18,2
C institution =>D institution, E institution	20	Request	%28,5	%7,01

There are 12 different subject groups in practice. The requested subject group is selected from Select CAtegory as shown in Figure 5.

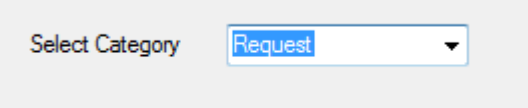

Figure5:Document category selection

The support value and the confidence value are taken as 1% and 22%, respectively, when the rules of association for the request group are obtained in Table 5. If we interpret these rules of association:

1. 25 documents sent to institution A and 8.77% of support value,
2. 10 documents sent to institution B and 3,5% of support value,
3. 70 documents sent to institution C and 24,5% of support value,
4. 109 documents sent to institution D and 38,2% of support value,
5. 65 documents sent to institution E and 22,8% of support value,
6. 5 documents sent to institution F and 1,75% of support value,
7. 8documents sent to institution G and 2,8% of support value,
8. 3documents sent to institution H and 1,05% of support value,
9. 30documents sent to institution J and 10,52% of support value,
10. 15 documents sent to institution K and 5,26% of support value,
11. 5 documents sent to institution L and 1,75% of support value,
12. 24% of documents coming to Institution A sent to Institution B and the support value was 2,1%,
13. 55,7% of documents coming to Institution C sent to Institution D and the support value was 13,6%,
14. 42,8% of documents coming to Institution C sent to Institution E and the support value was 10,52%,
15. 80% of documents coming to Institution Esent to Institution D and the support value was 18,2%,
16. We understand that 28.5% of documents coming to Institution C go to Institution D and Institution E and that the support value is 7.01%.

ISSN 2349-6193
Impact Factor: 2.805

IJESMR
International Journal OF Engineering Sciences &Management Research

From these results, the representative results can be derived as follows:

1. The largest number of documents, 109 documents and 38,2% of the support value went to D institution. From here, we can reach the result of the document that is in excess of the request of the institution D.
2. The paper with the lowest value of tabloda support is 3 documents and 1.05% support H institution. Different results can be obtained if the support value is reduced. Here the support value is 1%.
3. Among the two groups, the trust value is 80%, the support value is 18.2% and the highest with 52 documents is E institution and D institution. At the same time, we can say that most of the requests of the citizens are from these tables. Joint commissions can be established among these institutions. This will increase the performance of institutions and employees, as well as the faster and easier progress of citizens' affairs.
4. Among the three groups, the trust level is 28.5%, the support level is 7.01%, and the highest institutions are C institution, D institution and E institution with 20 documents. Looking at the rules of triple coexistence, 28.5% of documents coming to Institution C went to Institution D and Institution E as well. Joint commissions can also be created among these institutions.

We optained following results:

- How many documents have gone to which institutions
- The work intensity of institutions,
- Which institutions are more likely to collaborate with each other,
- We can gain the possibility that institutions can collaborate with each other by looking at their trust values [12].

RESULTS AND EVALUATION

In this paper, the apriori algorithm is used as one of the data mining methods. Developed this application, the rules of association are drawn according to a certain support value of the documents sent to the institutions. According to the conclusion of the rules of association, it has been determined which institutions should work together with which institutions in order to increase the performances of the institutions. It is aimed that the commissions can be established from various institutions and the performances of the institutions or employees with these commissions can be increased by these determinations. Moreover, thanks to these commissions, the work will be completed in a shorter period of time and the person will not have to travel between different institutions and will be able to complete their work in a single institution.

REFERENCES

1. *Chen, C. H., Lan, G. C., Hong, T. P., & Lin, S. B. (2016). Mining fuzzy temporal association rules by item lifespans. Applied Soft Computing, 41, 265-274.*
2. *Papamitsiou, Z.,& Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Journal of Educational Technology & Society, 17(4), 49.*
3. *Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3(6).*
4. *Bele, M. K. K.,& Phatak, A. A. (2016). Mining Association Rules Securely from Distributed System. Imperial Journal of Interdisciplinary Research, 2(5).*
5. *Moore, A. W.,&Zuev, D. (2005). Internet traffic classification using bayesian analysis techniques. In ACM Sıgmetrıcs Performance Evaluation Review (Vol. 33, No. 1, pp. 50-60). ACM.*
6. *Roiger, R. J. (2017). Data mining: a tutorial-based primer. CRC Press.*
7. *Singh, J., Ram, H., &Sodhi, D. J.(2013). Improving efficiency of apriori algorithm using transaction reduction. International Journal of Scientific and Research Publications, 3(1), 1-4.*
8. *Li, J., Sun, F., Hu, X., & Wei, W. (2015). A multi-GPU implementation of apriori algorithm for mining association rules in medical data. ICIC Express Letters, 9(5), 1303-1310.*
9. *Yang, Y., Gu, B., & Yoon, T. (2016). Deeper understanding of Flaviviruses including Zika virus by using Apriori Algorithm and Decision Tree. In MATEC Web of Conferences (Vol. 69, p. 01005). EDP Sciences.*
10. *Kim, H., Yoo, J., & Yoon, T. (2016). An Analysis of the Genomes of Dengue Virus Using Decision Tree and Apriori Algorithm. International Journal of Computer and Communication Engineering, 5(4), 294.*

ISSN 2349-6193
Impact Factor: 2.805

IJESMR
International Journal OF Engineering Sciences &Management Research

11. *Zhao, Y.,& Bhowmick, S. S. (2015). Association Rule Mining with R. A Survey Nanyang Technological University, Singapore.*
12. *Yıldırım, M. (2015). Harnessing Data Mining Techniques to Enhance the Performance in Interinstitutional Operations in a State.*