Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd

ISSN,e-ISSN: 2548-1304
Volume:7, Issue:2, pp:95-102 , 2022
https://doi.org/10.53070/bbd.1199055
Research Paper

Adaptation to Online Education: An Educational Data Mining Application

Çevrimiçi Eğitimde Adaptasyon: Bir Eğitsel Veri Madenciliği Uygulaması

Cengiz Hark*[1], Hatice Okumuş [2], Taner Uçkan[3]

[1] Bilgisayar Mühendisliği Bölümü, İnönü Üniversitesi, Malatya, Türkiye

[2]Malatya Yeşilyurt Bostanbaşı İlkokulu, Malatya, Türkiye

[3] Bilgisayar Mühendisliği Bölümü, İnönü Üniversitesi, Malatya, Türkiye

(cengiz.hark@inonu.edu.tr, tuthaticetut@gmail.com, taner.uckan@yyu.edu.tr)

Received:Nov.03,2022	Accepted:Nov.30,2022	Published:Dec.07,2022

Abstract— Despite space, time, and financial limitations, people who want to receive education participate intensively in online education programs that have emerged with the development of technology. With the Covid-19 outbreak, this interest has increased exponentially. In today's societies, where online education, which is preferred for different reasons, has become essential, examining the factors affecting success in online learning is a very important research topic. The study examined the level of adaptation to online education in terms of demographic variables. Experimental studies and necessary analyzes were carried out on the open-access 'Students Adaptability Level in Online Education' dataset. The results obtained using association rules, among the most widely used data mining techniques, have provided remarkable results regarding factors affecting success in distance education. It is thought that the study and the reported results will be a guide in creating education plans suitable for the demographic characteristics of the students enrolled in the online education program.
Keywords : Online education, data mining, association rules.

1. Introduction

We live in a rapidly changing world. Today, we need to benefit from technology in every field to keep up with the rapid change in science and technology. Technology; Today, it is in a continuous process of innovation and development in the fields of communication technology, educational technology, and information technologies. The ability to process, store and serve information quickly has made the computer the most sought-after tool in education. It has been widely accepted that computers should be used intensively in educational research, in the execution of educational services that have become more complex due to the increasing number of students, in student guidance-counseling activities and in the activities of measuring and evaluating success, and the applications have increased gradually. [1].

In our country and the developing world, people who want to receive education for different reasons, such as finding a job, making a career, and personal training, are enrolling in distance education programs that have recently increased their value in terms of number and quality. Distance education can be defined as a teaching service where instructors give lessons to students using internet technologies. Distance education is not only an academic service provided in public or private education institutions but also includes in-company training and community training provided by institutions and organizations to employees through distance education. Distance education programs have become an invaluable alternative to formal education for those who work or cannot participate continuously, with the flexibility of time in terms of courses and completion. [2].

Although the history of distance education dates back to the 1700s, its importance increased in 2020 for the first time in the history of the world, and the Covid-19 epidemic, which started towards the end of December 2019, has put distance education on the world agenda. When we look at the educational dimension of the epidemic around the world, it is seen that in countries where education is suspended due to Covid-19, institutions go to different practices and offer various resources for students, teachers, and parents within their means. In order not to interrupt the education and training process, the distance education process has started in Turkey, as in many countries. Face-to-face education has been suspended in all schools affiliated with the Ministry of National Education. While a holiday was announced for a few weeks in educational institutions at the beginning, it was foreseen that schools would not be able to be opened in the spring term, with the increase in the number of cases added to the death news. In order to maintain the education and training process, the distance education process has started in Turkey, as in many countries [3].

In education, data mining and analytics also have substantial transformative potential: they can be used to explore people's learning, predict learning, and understand actual learning behavior. Thus, a significant level of quality can be achieved in the outputs of the education system. Educational data mining can be used to achieve these goals, to design better and more innovative learning technology, and to better inform learners and educators [4]. Data mining studies in education provide access to data that can be useful to students, academics and educators in databases, based on the fact that there is information that has not yet been discovered [5].

In this study, the factors affecting success in distance education, which became widespread with the Covid-19 epidemic, were examined using association rules, one of the data mining tools. In addition, within the scope of the study, the level of adaptation to online education was examined in terms of demographic variables.

2. Educational Data Mining

Computer-based technologies have transformed our living habits, working, socializing, playing, and learning. Today, using data obtained with the help of these technological devices allows a second transformation in all these areas. In the last decade, data mining and data analytics methods have brought about significant changes in every field by discovering how people learn and understanding natural learning behavior [4][6].

Data mining is a research area that covers many disciplines. Education, astronomy, health, telecommunications, finance, banking, and marketing are the leading areas where data mining is used. The realization of data mining applications in education is called Educational Data Mining (EDM). By determining the characteristics, behaviors, and roles of individuals, EDM aims to provide services according to needs and to develop models according to these needs to ensure that individuals get better efficiency from the process they are in. It deals with the effective use of these models in decision-making and support processes. The main purpose of the EDM field is to examine how people learn depending on their psychology by making use of different disciplines (psychometrics, statistical techniques) and to use the recording data stored in offline education environments (learning management systems, intelligent teaching systems), including face-to-face contacts [7]. Akçapınar [8] stated in his study that learning analytics and EDM would be listed among the technologies that will guide educational research soon. He suggested that, due to the increase in the number of students daily, it will be crucial to investigate how to use these data effectively in online learning platforms and that studies in this direction should be focused on.

3. Association Rules

It is possible to examine data mining algorithms in three categories with a general evaluation. These algorithms are association rules, clustering and classification algorithms. In this study, algorithms based on association rules will be used.

Association rules, one of the data mining models, is used to find association behaviors among large data sets. These rules allow for the detection of unknown relationships and appropriate decision-making for more effective results [9] [10]. By revealing unknown relationships within data stacks, it provides results that will form the basis for decision-making and foresight [11]–[13].

Association rules try to determine the relationship between two movements by finding the probability of occurrence of B action (consequence) when an action A (precursor event) occurs. The rule support (s) of an association rule is the percentage ratio of the number of transactions involving $A \cup B$ to the total number of transactions (N) in the database. Rule support is formulated as [1], [3]–[5], [7], [14].

$$\text{Support } = \frac{\text{n(X } \cup \text{ Y)}}{N} \tag{1}$$

The expression $n(X \cup Y)$ in the formula represents the number of transactions where X and Y are together, and N represents the total number of transactions.

Confidence value, on the other hand, expresses what percentage of transactions containing X also includes Y and is calculated with the help of Formula 2:

$$\text{Confidence } = \frac{\text{n(X } \cup \text{ Y)}}{n(X)} \tag{2}$$

Unlike the support formula, there are a number of observations containing the total X in the denominator.

A support value of one indicates that X and Y occur together in every transaction in the analyzed data set. A zero indicates that X and Y do not occur together in any transaction in the data set. A confidence value of one indicates that every transaction containing X also includes Y, and a zero indicates that none of the transactions involving X contain Y.

For the implementation of the Apriori algorithm, the minimum confidence and minimum support values determined by the user must be determined.

3.1. Apriori Algorithm

The Apriori algorithm, a simple and well-known algorithm that extracts association rules from data sets, has been the most applied algorithm for extracting association rules in the history of data mining [15]. The Apriori algorithm enables the extraction of association rules with support and confidence above a specified threshold. It has a repetitive nature, and it is necessary to scan the database many times to find the datasets that frequently occur in the databases [5].

Association rules with support and trust above the threshold are called strong rules. In some sources, these rules are also called interesting rules. It is possible to generate many rules in a medium-sized dataset. However, it is a fact that the rules that do not cover the whole data set are not worth examining. For this reason, it is necessary to list only those rules that meet certain conditions, not every rule. It is possible to say that the minimum trust and minimum support values limit the number of rules. The highest support and confidence value is 100% (or 1), indicating that the relevant rule is included in the entire data set. The Apriori algorithm can provide useful information for marketing. At the same time, the outputs are easy to read. Despite these advantages, the rules can be difficult to understand when many rules are removed. The second disadvantage of the algorithm is that it needs extremely long computation time in large data sets [15], [16].

Algorithm 1. Pseudocode of the Apriori algorithm

```
L1= { frequent item set };
for (k=2; Lk-1≠ Ø; k++) do begin
    Ck=apriori-gen (Lk-1); // new candidates
    forall transactions-hareketler t ∈ D do begin
        Ct = subset (Ck, t); // Candidates are in t
        forall candidates – adaylar c ∈ Ct do
            c.count++;
    end
    Lk = {c ∈ Ck | c.count ≥ minsup}
end
Answer = ∪k Lk;
```

The pseudo-code of the Apriori algorithm is seen in Algortima 1 [17].

When the Apriori algorithm's summary code is examined, the database is scanned many times to find frequent item sets. Before the first step, the data collection that will be used for data mining is scanned to determine how many of the items are included in the motion records (this value determined for each item is called the support counter) and the items whose support counter is equal to or greater than the minimum support value are L1 frequent 1-item. Assuming that the set is determined as a set, the process starts. With the loop structure established in the code, a new set is formed in a way similar to the binary combination of the elements of the L1 frequent item set ($L1 \infty L1$) in the first stage, and this process is called a join. The clusters formed by this process are also called candidate item clusters and are symbolized by the letter C. Since each element of this candidate element set consists of two elements, it is named with the expression C2. Next, this candidate set is pruned with the apriori-gene function, and it is checked whether the subsets of the elements of the C2 set are in the L1 item set. Scanning the data collection with the Apriori algorithm shows how many motions record the C2 candidate set elements passed through the pruning process (support counter). By the support counter information found, the elements with support counter equal to or greater than the minimum support value of the C2 candidate set elements to form the L2 frequent item set. In the next step, the loop creates a new candidate item set with the triple combination of the elements of the set L2, and the expression C3 represents this set. As in the first stage, this cluster goes through the pruning process, and after pruning, the L3 frequent item set is created with the elements above the minimum support level. The loop continues, increasing the number of items on each turn. This process continues until a new set of frequent items cannot be found [5], [15]–[17].

4. Application Steps

Within the scope of the study, the level of adaptation to online education was examined in terms of demographic variables. Experimental studies and necessary analyzes were carried out on the "Students Adaptability Level in Online Education" dataset, which has open access. Association rules, which are among the most widely used data mining techniques, were used.

4.1. Data Set

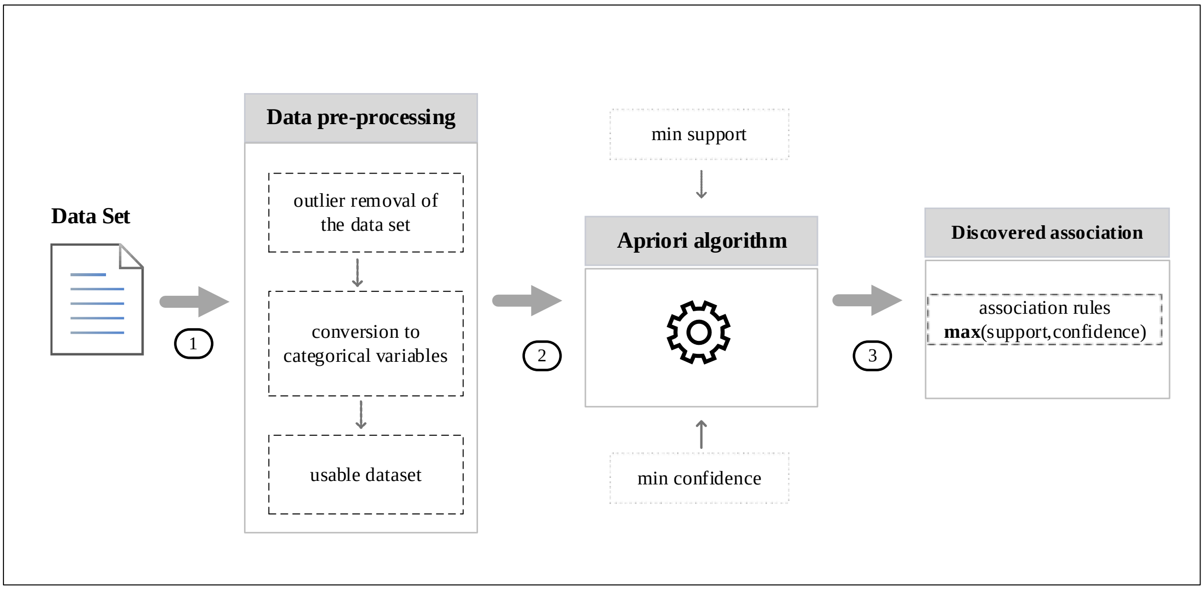
The researcher group collected the data used in this study through a questionnaire [18]. Students with different education levels participated in the survey. In the data set, which includes 1205 data, 14 various features are included. Detailed information about the data set is given in Table 1.

Table 1. Variable details with possible values for the Data Set

Attribute	Type	Possible Values
Gender	Independent	{Boy, Girl}
Age	Independent	{1-5,6-10,11-15,16-20,21-25,26-30}
Education Level	Independent	{University,College,School}
Institution Type	Independent	{Government,'Non Government'}
IT Student	Independent	{'Yes', No'}
Location	Independent	{Yes', 'No'}
Load-shedding	Independent	{Low', 'High' }
Financial Condition	Independent	{'Mid', Poor, Rich}
Internet Type	Independent	{'Mobile Data', Wifi}
Device	Independent	{'Mobile', 'Tab', 'Computer'}
Network Type	Independent	{2G, 3G, 4G}
Class Duration	Independent	{0,'1-3','3-6'}
Self Lms	Independent	{Yes, No}
Adaptivity Level	**Dependent**	{'Moderate','Low','High'}

4.2. Application Steps

Experimental studies and necessary analyzes were carried out on the open-access “Students Adaptability Level in Online Education” data set. Since the researchers previously obtained the data set, a process for recovering the data set was not carried out. During the data cleaning phase, the records in the database containing mostly missing data were identified and deleted. In addition to this, certain processes were followed by adhering to data integrity, and missing areas were completed. After this stage, it has come to the stage of determining the modeling and algorithm to obtain the golden data from the existing data. This stage is very important. Some association rule extraction methods were tried on the data set used, and it was decided to use the Apriori algorithm due to its effective results. Various minimum support and minimum confidence values were tried. In the last stage, the striking relationships and associations between the independent variables and the dependent variable in the data set were revealed.



4.3. Experimental Results

Within the scope of the study, experimental studies were carried out on a very popular topic in today's societies, such as determining the adaptation levels of students with different demographic characteristics towards online learning. In this context, the level of adaptation to online education was examined in terms of demographic variables. Experimental studies and necessary analyzes were carried out on the "Students Adaptability Level in Online Education" dataset, which has open access. Association rules, which are among the most widely used data mining techniques, were used. The results provided remarkable results regarding demographic factors affecting success in distance education.

Among the many rules obtained as a result of the experimental processes, selections were made in a way that had high confidence values and was compatible with the scope of the research. Table 2 shows the rules that reveal the adaptations of demographic characteristics to online learning with the association rule method. In the relevant table, the demographic characteristics whose effects were investigated due to the purpose and scope of the study were brought together and included in the table by grouping. This step is to present the effect of demographic characteristics on adaptation levels in a more effective and understandable way.

Table 2. Some association rules obtained for adaptation at the end of the experimental process

Demographic Attributes		Association Rules	Conf(%)
1	**Frequency of Power Outages**	Non it, **Having frequent power cuts**, Lesson attendance time is 0 ==> Low Adaptation	100
		Living in rural area, **Having frequent power cuts**, Lesson attendance time is 0 ==> Low Adaptation	100
		Attending private school, Non it, Living in the city center, **Rarely experiencing power outages,** Middle-income, Using mobile network, Class attendance time is 1-3 ==> It is moderately adaptable	85
		Education level university, Living in the city center, **Having frequent power cuts**, Middle-income, Using mobile network ==> It is moderately adaptable	100
2	**Lesson Attendance Time**	Non it, **Lesson attendance time is 0**, Mobile device user==> Low Adaptation	94
		Non it, **Lesson attendance time is 0**, No corporate LMS, Mobile device user 170 ==> Low Adaptation	94
		Lesson attendance time is 0, Mobile device user ==> Low Adaptation	93
		Lesson attendance time is 0, No corporate LMS, Mobile device user==> Low Adaptation	93

3	**No corporate LMS**	Woman, Non it, Lesson attendance time is 0, **No corporate LMS** ==> Low Adaptation conf:(1)	100
		Woman, Lesson attendance time is 0, **No corporate LMS**, Mobile device user ==> Low Adaptation	100
		Woman, Non it, Lesson attendance time is 0, **No corporate LMS**, Mobile device user 80 ==> Low Adaptation	100
		Woman, Aged between 11-15, Attending private school, Living in the city center, Rarely experiencing power outages, Middle-income, **No corporate LMS**==> It is moderately adaptable	100
		Woman, Aged between 11-15, Education level School, Attending private school, Living in the city center, Rarely experiencing power outages, Middle-income, **No corporate LMS** ==> It is moderately adaptable	100
4	**Location**	Non it, **Living in the city center**, Rarely experiencing power outages, Lesson attendance time is 0==> Low Adaptation	100
		Woman, Aged between 11-15, **Living in the city center**, Rarely experiencing power outages, Network Tipi 4G olan, No corporate LMS==> It is moderately adaptable	100
		Woman, Aged between 11-15, **Living in the city center**, Rarely experiencing power outages, Class attendance time is 1-3, No corporate LMS==> It is moderately adaptable	100
5	**Age Group**	**aged 16-20 years**, Eğitim seviyesi Kolej olan, Living in rural area, Rarely experiencing power outages, Network Tipi 4G olan, No corporate LMS, Mobile device user 50 ==> Low Adaptation	100
		Aged between 11-15, Education level School, Non it, Lesson attendance time is 0==> Low Adaptation	100
		Woman, **Yaşı 26-30 aralığında olan**, Devlet Okuluna giden, Network Tipi 4G olan==> Low Adaptation	100
		Woman, **Yaşı 21-25 aralığında olan**, Devlet Okuluna giden, Non it, No corporate LMS, Mobile device user ==> Low Adaptation	100
6	**Institution Type**	Education level School, **Attending private school**, Using mobile network, Network Tipi 4G olan ==> It is moderately adaptable	84
		Education level School, **Attending private school**, Living in the city center, Using mobile network, Network Tipi 4G olan==> It is moderately adaptable	84
		Education level university, **Devlet okuluna giden**, Living in the city center, Middle-income, Network Tipi 4G olan, Lesson attendance time is 0 ==> Low Adaptation	100
		Education level School, **Attending private school**, Non it, Living in the city center, Lesson attendance time is 0, No corporate LMS ==> Low Adaptation	100
		aged 16-20 years, Education level College, **public school going**, Living in rural area, Rarely experiencing power outages, Network Tipi 4G olan, No corporate LMS ==> Low Adaptation	100

However, a selection was made considering the rules shared in Table 2, among the many rules obtained as a result of the analysis, considering the association rules with high confidence values and which can be considered significant according to the scope of the research. The important results of this selection are presented in the relevant table.

5. Discussion and Conclusion

The reported high confidence values are essential in expressing how strong the obtained association rules are.

When the obtained rules are examined, the level of adaptation needs to be improved in the rules that include the information that power cuts are frequently experienced. However, it is observed that mobile internet users are less affected by the mentioned power cuts. It may be possible to explain this result because the related companies have reached wider coverage areas with their developing mobile network infrastructures.

It is seen that the duration of the lesson has a direct and indisputable effect on adaptation. When the association rules reported in the related table are examined, it is seen that the adaptation levels are also quite low in the rules where the attendance time is low. This finding indicates that the duration of class participation increases the level of adaptation.

It is understood from the many different association rules with different demographic characteristics that whether there is an LMS system belonging to the Institution or not is not effective on student adaptation levels. Because when the association rules included in the related tables are examined, many rules with low, medium, and high adaptation levels and high confidence values have been formed. These rules are not affected by the absence of an LMS infrastructure belonging to the Institution.

Similarly, it is seen that the participants' residence in the city center or a rural area does not affect the level of student adaptation.

Similarly, the distinction of being educated in a public or private school did not create a significant difference in adaptation levels.

It was observed that the different age groups in the data set used did not show any difference in terms of student adaptation levels. The reason why age does not make a significant difference in the level of adaptation to the increasing frequency of technology use in each age group can be presented.

The study and the reported results can guide the creation of education plans suitable for the demographic characteristics of the students enrolled in the online education program.

6. Resources

[1] N. H. SÖYLEMEZ, “BİLGİSAYAR DESTEKLİ VE BİLGİSAYAR TEMELLİ ÖĞRETİM YÖNTEMLERİNİN AKADEMİK BAŞARI VE KALICILIĞA ETKİSİ,” *Elektron. Eğitim Bilim. Derg.*, vol. 2, no. 3, Mar. 2013, Accessed: Oct. 17, 2022. [Online]. Available: https://dergipark.org.tr/tr/pub/ejedus/issue/15937/167581

[2] M. Aksarayli, D. İibf, and İ. / Türkiye, “Uzaktan Eğitimi Tercih Etme Nedenleri ve Başarı Arasındaki İlişkinin Kümeleme Analizi İle İncelenmesi,” *West. Anatolia J. Educ. Sci.*, vol. 8, no. 2, pp. 37–48, Dec. 2017, Accessed: Oct. 14, 2022. [Online]. Available: https://dergipark.org.tr/en/pub/baebd/issue/33149/350179

[3] S. G. TELLİ and D. ALTUN, “Coronavirüs ve Çevrimiçi (Online) Eğitimin Önlenemeyen Yükselişi,” *Üniversite Araştırmaları Derg.*, vol. 3, no. 1, pp. 25–34, Apr. 2020, doi: 10.32329/uad.711110.

[4] R. S. Baker, “Educational data mining: An advance for intelligent systems in education,” *IEEE Intell. Syst.*, vol. 29, no. 3, pp. 78–82, 2014, doi: 10.1109/MIS.2014.42.

[5] C. Hark, “Öğrencilerin Akıllı Tahtaya İlişkin Tutumlarının İncelenmesine Yönelik Bir Veri Madenciliği Uygulaması. Fırat Üniversitesi Eğitim Bilimleri Enstitüsü, Yüksek Lisans Tezi,” Fırat Üniversitesi, 2013.

[6] T. UÇKAN, C. HARK, and A. KARCI, “Fp Growth Algoritması Ve Big Data Uygulamaları,” in *IDAP 2016 - International Artificial Intelligence and Data Processing Symposium*, 2016, pp. 338–341.

[7] G. ÇETİNTAV, B. D. ÇİL, and R. YILMAZ, “EĞİTSEL VERİ MADENCİLİĞİ VE ÖĞRENME ANALİTİKLERİ ARAŞTIRMALARINDA VERİ GİZLİLİĞİ VE ETİK MESELELER: ARAŞTIRMALAR ÜZERİNE BİR İNCELEME,” *Eğitim Teknol. Kuram ve Uygul.*, vol. 12, no. 1, pp. 113–146, Jan. 2022, doi: 10.17943/ETKU.950392.

[8] Maciej Serda *et al.*, “Çevrimiçi Öğrenme Ortamındaki Etkileşim Verilerine Göre Öğrencilerin Akademik Performanslarının Veri Madenciliği Yaklaşımı İle Modellenmesi,” *Uniw. śląski*, vol. 7, no. 1, pp. 343–354, 2014, doi: 10.2/JQUERY.MIN.JS.

[9] C. HARK, T. UÇKAN, and A. KARCI, “15 Temmuz Şehitler Köprüsü Kapsamında Hava Yol Durum Bilgileri Kullanılarak Trafik Kazalarının Nedenlerine Yönelik Birliktelik Çıkarımı,” in *IDAP 2016 -*

International Artificial Intelligence and Data Processing Symposium, 2016, pp. 347–351.

[10] C. W. Liao and Y. H. Perng, “Data mining for occupational injuries in the Taiwan construction industry,” *Saf. Sci.*, vol. 46, no. 7, pp. 1091–1102, Aug. 2008, doi: 10.1016/J.SSCI.2007.04.007.

[11] G. Mühendislik *et al.*, “TRAFİK KAZALARININ BİRLİKTELİK KURALLARI İLE ANALİZİ,” *Gazi J. Eng. Sci.*, vol. 1, no. 2, pp. 265–283, Jun. 2015, Accessed: Oct. 20, 2022. [Online]. Available: https://dergipark.org.tr/en/pub/gmbd/issue/29232/312922

[12] M. Aydoğan and A. KARCI, “Meslek Yüksekokulu Öğrencilerinin Başarı Performanslarının Makine Öğrenmesi Yöntemleri ile Analizi,” in *2nd International Symposium on Multidisciplinary Studies and Innovative Technologies*, 2018, pp. 1–5.

[13] M. Aydogan and A. Karci, “Turkish Text Classification with Machine Learning and Transfer Learning,” *2019 Int. Conf. Artif. Intell. Data Process. Symp. IDAP 2019*, Sep. 2019, doi: 10.1109/IDAP.2019.8875919.

[14] A. MAKALESİ Ayfer ALPER and A. Üniversitesi Eğitim Bilimleri Fakültesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, “PANDEMİ SÜRECİNDE K-12 DÜZEYİNDE UZAKTAN EĞİTİM: DURUM ÇALIŞMASI,” *Milli Eğitim Derg.*, vol. 49, no. 1, pp. 45–67, Dec. 2020, doi: 10.37669/MILLIEGITIM.787735.

[15] İ. Söylemez *et al.*, “Association Rules on Traffic Accident: Case Of Ankara,” *Ege Acad. Rev.*, vol. 16, no. 5, pp. 11–20, Nov. 2016, doi: 10.7240/MUFBED.56489.

[16] V. Madenciliğinde, B. Kuralları, İ. El, O. Piyasası, and Ü. Bir Uygulama, “Veri Madenciliğinde Birliktelik Kuralları ve İkinci El Otomobil Piyasası Üzerine Bir Uygulama,” *Ordu Üniversitesi Sos. Bilim. Enstitüsü Sos. Bilim. Araştırmaları Derg.*, vol. 7, no. 1, pp. 45–58, Mar. 2017, Accessed: Oct. 20, 2022. [Online]. Available: https://dergipark.org.tr/tr/pub/odusobiad/issue/28361/301580

[17] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules in Large Databases,” in *VLDB’94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*, 1994, pp. 487–499. [Online]. Available: http://www.vldb.org/conf/1994/P487.PDF

[18] M. M. H. Suzan, N. A. Samrin, A. A. Biswas, and M. A. Pramanik, “Students’ Adaptability Level Prediction in Online Education using Machine Learning Approaches,” *2021 12th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2021*, 2021, doi: 10.1109/ICCCNT51525.2021.9579741.