Fırat Üniversitesi Müh. Bil. Dergisi
36(1), 301-310, 2024

Araştırma Makalesi
https://doi.org/10.35234/fumbd.1332199

Evaluation of Academic Self-Efficiency, Community Feeling, and Academic Achievement of Students in the Process of the Covid-19 Pandemic by Data Mining Techniques

Songül KARABATAK[1*], Özal YILDIRIM[2], Murat KARABATAK[3]
[1]Eğitim Yönetimi, Eğitim Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye
[2,3]Yazılım Mühendisliği, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye
[1]skarabatak@firat.edu.tr, [2]oyildirim@munzur.edu.tr, [3]mkarabatak@firat.edu.tr

(Geliş/Received: 24/07/2023; Kabul/Accepted:19/03/2024)

Abstract: Thanks to the advancement of technology, vast amounts of data are being generated in various fields on a daily basis. The research on identifying hidden patterns and extracting useful information from big data has become increasingly important. In the field of education, the availability of large datasets has allowed for the emergence of data mining techniques as an alternative to traditional statistical methods. Unlike traditional statistical methods, data mining can uncover hidden relationships between variables, thus avoiding the loss of valuable information and enabling the utilization of essential data in education. By unlocking valuable insights and predicting important relationships, educational data mining (EDM) has the potential to enhance and improve the quality of education. This study aims to demonstrate the predictive power of EDM through a sample application and draw attention to its implications. The dataset used in this study consists of survey responses collected from university students. The variables in the dataset include academic self-efficacy, sense of community, academic achievement averages, and various demographic variables of distance education students. Descriptive modeling was employed to identify latent patterns between variables, while a predictive model was utilized to estimate variables. In order to achieve this, both association rule mining and classification algorithms were employed. The findings of this study indicate that EDM can effectively identify relationships between variables and make accurate predictions.

Key words: Educational data mining, academic self-efficiency, community feeling, academic achievement.

Covid-19 Pandemisi Sürecinde Öğrencilerin Akademik Öz Yeterlilik, Topluluk Hissi ve Akademik Başarılarının Veri Madenciliği Teknikleri ile Değerlendirilmesi

Öz: Günlük hayatın bir parçası haline gelen teknolojiler sayesinde hemen her alanda devasa veri yığınları oluşmaktadır. Büyük verideki gizli örüntülerin tespit edilmesi ve faydalı bilgilerin keşfedilmesine yönelik araştırmalar önem kazanmıştır. Eğitim alanında biriken veri miktarı, bu alanda geleneksel istatistiksel yöntemlere alternatif olarak veri madenciliği tekniklerinin ön plana çıkmasını sağlamıştır. Geleneksel istatistiksel yöntemlerde bazı değişkenler arasındaki gizli ilişkiler göz ardı edilebilmektedir. Bu da bazı bilgilerin kaybolmasına ya da eğitim gibi temel alanlarda gerekli verilerin kullanılamamasına neden olabiliyor. Ancak eğitimsel veri madenciliği (EVM), eğitimin kalitesini iyileştirmek ve geliştirmek için değerli verilerin kilidini açabilir ve önemli ilişkileri tahmin edebilir. Bu nedenle bu çalışma, EVM'nin tahmin gücüne dikkat çekmek için örnek bir EVM uygulaması gerçekleştirmeyi amaçlamıştır. Veri seti üniversite öğrencilerinden toplanan görüşlerden oluşmaktadır. Bu veri setinin değişkenlerini uzaktan eğitim öğrencilerinin akademik öz yeterlilikleri, topluluk hissi, akademik başarı ortalamaları ve bazı demografik değişkenler oluşturmuştur. Betimsel model, çalışmadaki değişkenler arasındaki örtük örüntüleri ortaya çıkarmış ve değişkenleri tahmin etmek için yordayıcı bir model kullanılmıştır. Bunun için birliktelik kuralı yöntemi ve sınıflandırma algoritması da kullanılmıştır. Çalışma sonunda EVM'nin değişkenler arasındaki ilişkileri etkili bir şekilde bulabildiği ve değişkenleri tahmin edebildiği sonucuna varılmıştır.

Anahtar kelimeler: Eğitsel veri madenciliği, akademik öz yeterlilik, topluluk hissi, akademik başarı.

1. Introduction

The products of extensive efforts to extract valuable information from data have started to surface. These studies, coupled with advancements in software technologies, have led to the development of numerous methods. These methods have diverged from traditional statistical approaches in certain aspects [1]. Among these innovative methods is data mining, which has gained popularity as a data analysis technique in recent times.

Data mining is the process of extracting implicit, meaningful, and valuable information from large datasets [2,3]. Similarly, data mining is defined as the task of obtaining 'valuable' information from large-scale data. This method enables the revelation of relationships between the data and facilitates future predictions when necessary

[*] Sorumlu yazar: skarabatak@firat.edu.tr. Yazarların ORCID Numarası: [1] 0000-0002-1303-2429, [2] 0000-0001-5375-3012, [3] 0000-0002-6719-7421

[4]. Data mining is used to uncover data patterns, organize data regarding hidden relationships, construct association rules, estimate values of unknown items to classify objects, create clusters of homogeneous objects, and uncover various types of findings that cannot be quickly produced by a classical computer-based information system [5,6].

Mathematicians conducted the first studies on data mining in logic and computer science in the 1950s. These studies later transitioned to artificial intelligence and machine learning technologies and found applications in various fields such as commerce, tourism, medicine, insurance, communication, production, biology, bioinformatics, and social sciences [7,8]. Fundamentally, data mining models can be categorized into two types [5]: descriptive and predictive models. Descriptive models reveal patterns, relationships, or correlations within the data [9]. Descriptive models often utilize unsupervised learning techniques to generate designs that capture the underlying structure and relationships within the data [10]. Descriptive data mining methods uncover the general characteristics of the dataset [11]. In descriptive models, existing data patterns that can guide decision-making are defined. The analysis results provide an opportunity to understand the relationships present in the dataset without any prior hypothesis [12].

Predictive models, on the other hand, create one or more datasets, make inferences from the existing dataset, and attempt to predict the behavior of new datasets [9]. Predictive models often employ supervised learning techniques to predict unknown or future values of dependent variables based on the properties of relevant independent variables [13]. Predictive models play a crucial role in decision-making processes. They aim to develop a model based on data with known outcomes and use this model to estimate the outcomes for datasets with unknown results [12].

The development of data mining also enables the processing of training data [14]. Data mining methods are also utilized to discover unique types of data originating from educational environments and gain a better understanding of individuals' learning or skills [15]. Descriptive and predictive models of data mining are also employed to analyze training data. Educational data mining (EDM) can be defined as the application of data mining methods to training data [16]. In other words, EDM can be considered as a data mining method that focuses on educational content [17].

EDM's primary objective is to utilize large-scale educational datasets to gain a deeper understanding of learning and provide valuable insights into the learning process [18]. Therefore, EDM relies on data mining to explore data from educational settings and uncover descriptive patterns and predictions that characterize learners' behaviors and achievements, domain knowledge content, assessments, educational functionalities, and applications [5]. EDM encompasses a set of processes that extract meaningful information from extensive and rich educational datasets using methods such as prediction and classification [19].

The utilization of EDM has experienced a substantial increase due to its focus on developing strategies for analyzing the unique types of data found in an academic context. By doing so, EDM offers new avenues for addressing long-standing research problems in traditional educational technology. Consequently, it emerges as a valuable resource for predicting and enhancing the quality of education [20]. Given the challenges associated with analyzing current learning behaviors using traditional research methods, EDM stands as one of the most practical applications for predicting learner performance [21]. The proliferation of e-learning systems has led to an increase in the use of diverse assessment methods. EDM has the potential to fulfill this growing need and further expand educational possibilities [22]. Moreover, the student-centered nature of education and the demand for personalized instruction underscore the importance of utilizing data mining techniques on educational data [5]. By employing EDM methods, it becomes feasible to analyze student behaviors effectively, which would otherwise remain elusive using traditional research methods [21].

EDM-based assessments are widely recognized for their numerous advantages over traditional assessment methods [22]. When appropriately designed, these assessment models allow for evaluation during the learning process, instant intervention, minimizing time loss, and easy repetition of assessments conducted with traditional tools. By leveraging the ongoing interaction between students and online systems, these models can replicate assessments without the need for time-consuming paper tests. Furthermore, EDM serves as a valuable resource for predicting and improving the quality of education. It unveils student achievements and provides various suggestions for enhancing success [23]. Some researchers argue that the challenges faced in higher education result from a knowledge gap caused by insufficient training processes such as planning, evaluation, and consultancy. They propose using data mining-based models to address these issues [24]. By leveraging data mining, the hidden patterns, relationships, and anomalies can be uncovered, thus bridging the knowledge gap [25]. Additionally, EDM enables the interpretation of students' perspectives [5].

EDM can extract meaningful and valuable insights from various data sources within educational systems, including gender, age, motivation, emotional states, exam grades, and attitudes [18, 26, 27]. The literature highlights several advantages of EDM, including fast and meaningful results, easy applicability, and more efficient identification of individual differences compared to traditional methods [28]. Furthermore, data mining can generate meaningful and valuable information from online and distance education data.

Digital platforms such as online and distance education environments, which were popular prior to the pandemic, have now become the prevailing norm within the education system as a result of the pandemic and the need for learning management systems (LMS). As learners engage with these systems in various ways, they generate extensive data in the form of electronic traces. This wealth of data has been recognized as a valuable resource for understanding the learning process and outcomes [29]. Consequently, the data stored within LMS has gained significant value, but it needs to be effectively analyzed. In other words, while the collection of data from LMS learners has become easier, the analysis of digital educational information has become more complex [21]. Consequently, there is a high demand for research based on educational data mining (EDM) [29], as online systems store data that can be analyzed to address numerous issues [30].

EDM is also utilized to assess students' performance based on different needs and cognitive characteristics, as well as to understand their learning styles [3, 27, 31]. This capacity has the potential to significantly guide the use of student data accumulated within LMS, enhancing the effectiveness of distance education processes and facilitating personalized instruction. Rodrigues et al. [32] also emphasized in their study that EDM methods should be employed to enhance the teaching and learning process for students. Therefore, the objective of this study is to conduct a sample EDM application to highlight its predictive capabilities. The dataset for this application consists of the opinions of university students who participated in distance education during the COVID-19 pandemic. It was hypothesized that the pandemic would impact students' academic self-efficacy, sense of community, and academic achievements. Based on this assumption, this study aims to uncover and predict the hidden, significant, and meaningful relationships between demographic characteristics (age, gender, and department), academic achievements, academic self-efficacy, and community feelings of students engaged in distance education through the utilization of data mining methods.

2. Method and Material

In this study, a relational survey model was employed. The relational survey model elucidates the rationale for enhanced modeling by examining the potential correlation between two or more variables [33,34]. Additionally, predictive and descriptive data mining techniques were utilized to unveil latent patterns. Figure 1 illustrates the schematic diagram of the methodology employed in the study.

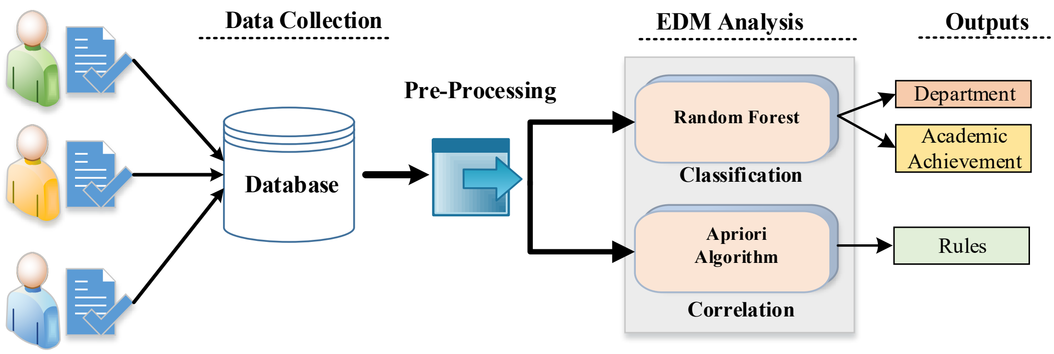

Figure 1. A schematic diagram illustrating the methodologies employed in the analysis

2.1. Education Data Set

The study was conducted during the spring semester of the 2019-2020 academic year. The research group consisted of 1898 students enrolled in the education faculty of a university located in a province in eastern Turkey. The survey was distributed electronically to the entire group via the distance education system. A total of 467 students from this population completed the form. However, 84 forms with missing data were excluded from the dataset. The final dataset included 383 student responses. Of these participants, 231 were female and 152 were male. The age range of participants was between 19 and 41. Some students (n=80) reported using the internet for a short duration (0-3 hours), while others (n=190) used it moderately (3-7 hours), and some (n=113) used it extensively (more than 7 hours). In terms of accessing course documents, 89 students reported experiencing difficulties, while the majority (n=294) stated that they did not encounter any issues. The distribution of students across departments who participated in the study is shown in Figure 2.

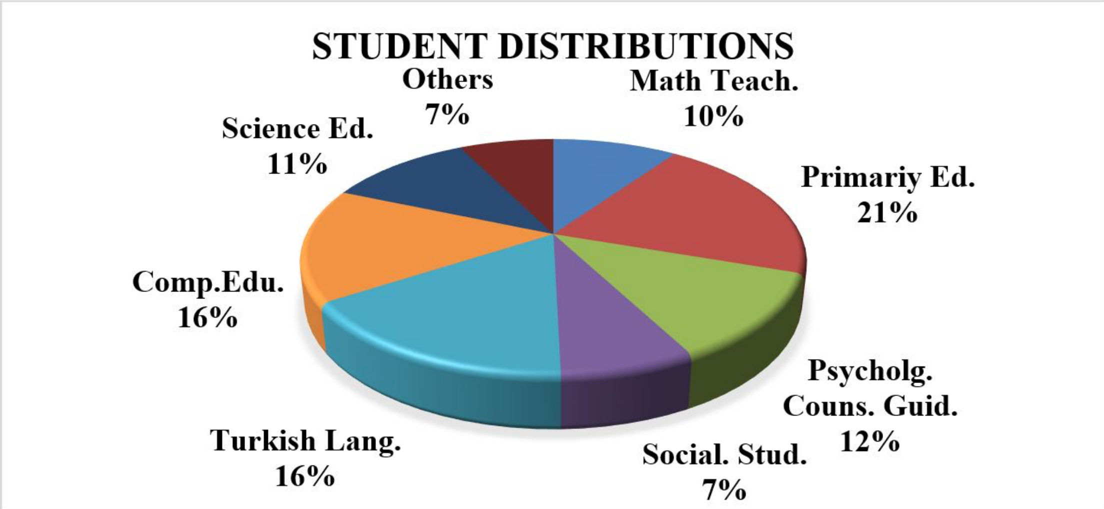

Figure 2. The department distribution rates of the students

As depicted in Figure 2, students are enrolled in eight distinct departments: Mathematics Teaching (n=37), Primary Education (n=79), Psychological Counseling and Guidance (n=46), Social Studies Teaching (n=28), Turkish Language Teaching (n=62), Computer Education Teaching (n=60), Science Teaching (n=43), and other departments (n=28).

2.2. Data Collection Tools

The data collection tool used consists of two parts. First, the student's demographic characteristics (gender, department, and age) and internet usage status (daily internet usage duration and difficulty in accessing the course documents). The second part included academic achievement (GPA of the previous distance education term), Academic Self-Efficacy Scale, and Community Feelings Scale.

The Academic Self-Efficacy Scale was developed in 1981 by Jerusalem and Schwarzer [35] and adapted into Turkish by Yilmaz, Gürçay, and Ekici [36]. The scale consists of seven items. The Cronbach's alpha reliability value of the original scale was .87. The Adapted Turkish scale Cronbach Alpha reliability value was determined as .79. In the present study, the reliability coefficient was calculated as .86. *The Community Feelings Scale* was developed by Rovai et al. [37] and adapted into Turkish by Ilgaz and Askar [38]. The reliability coefficient of the scale, Cronbach alpha coefficient, was found to be .80. In this study, the reliability value of the scale was calculated as .83. The items of the Academic Self-Efficacy Scale and the Community Feelings Scale are given in Appendix 1.

2.3. Data Pre-Processing Stage

One of the most critical steps in the data mining process is the pre-processing stage. The success of the pre-processing stage is vital in obtaining accurate and effective results. During the pre-processing stage, the data set was first filtered to eliminate noisy data (n=87). The data collected with the Academic Self-Efficacy Scale were categorized as Sufficient, Partially Sufficient, and Insufficient. The Community Feelings Scale data were grouped as *Low, Medium,* and *High*. The daily internet usage duration data were also grouped as *Little, Moderate,* and *Lot*. The difficulty accessing the course documents was categorized as Yes and No. All academic achievement averages were converted into a *4-point system.*

2.4. The Proposed Data Mining Techniques

There are several methods and algorithms used in data mining for converting data into information. These methods include classification, clustering, predictive modeling, data visualization, change and deviation detection analysis, and association rules [39]. The choice of algorithm depends on the method used for data grouping. In this study, the focus is on classification algorithms as an example of a predictive model, and association rules as an example of a descriptive model.

Classification is the process of appropriately assigning data to defined classes within a dataset. Classification algorithms, on the other hand, are used to learn the characteristics of different types using educational data and predict which category new information belongs to [40]. The most commonly used classification algorithms in the literature are Naive Bayes, Artificial Neural Networks (ANN), Support Vector Machines (SVM), K-Nearest Neighbor (IBK), KStar, C4.5 (J48), and Random Forest algorithms. In this study, all algorithms were tested in the classification process, and it was found that the Random Forest Algorithm produced the best results. Therefore, the results of the Random Forest algorithm are presented in this study. The Random Forest Algorithm is a tree-based classification algorithm that generates multiple classifiers and classifies new data based on the estimation results of these classifiers [41].

To evaluate the models created using classification algorithms and determine which classification model produces more accurate results, certain evaluation metrics are used. These metrics are based on a table known as the confusion matrix. Each row of the matrix represents the actual values, while each column represents the predicted values [42].

The Association Rule focuses on the frequency of collected data and the relationships between items. To make relationship inferences, association rules must meet specific metrics. These metrics include *support, confidence,* and *lift.* The *support value* indicates the number of occurrences of groups A and B together in the dataset. The *confidence value* is the frequency of finding group B in the relationships where group A is present in the dataset. The *lift value* represents the frequency of co-occurrence of groups A and B when they are independent [43]. In this study, the term "group" is specified as the *variable*.

Apriori, segmentation, sampling, AIS, SETM, DIC, and CHARM are algorithms used to discover relationships in association rules [23]. The Apriori algorithm is an improvement over the AIS and SETM algorithms. The Apriori algorithm was chosen in this study due to its ease of processing, speed, and accuracy compared to other algorithms. Figure 3 provides a representation of how the Apriori algorithm operates.

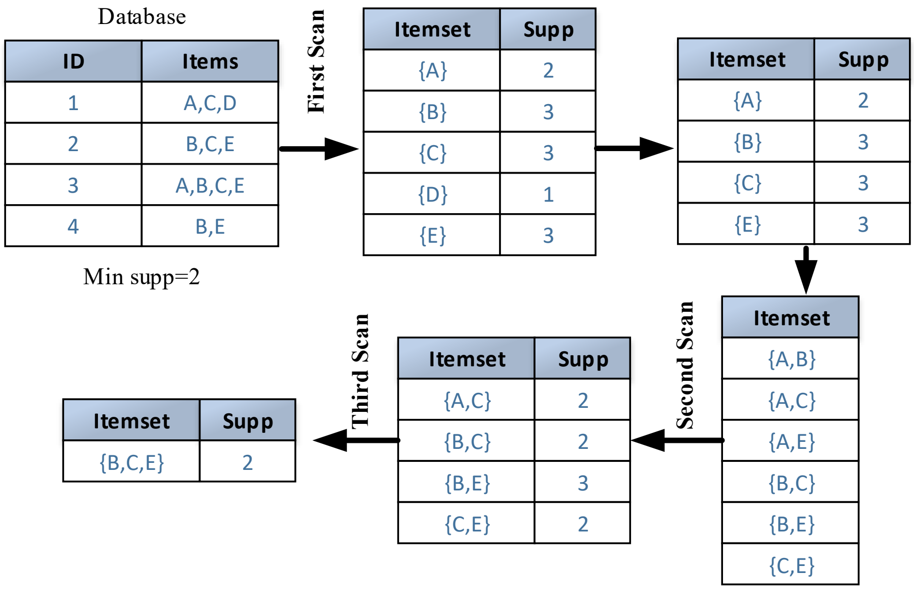

Figure 3. A representation of the operation of the Apriori algorithm

3. Experimental Results

The findings regarding the classification performed to determine the predictive accuracy of variables (department and academic achievement) and the association rule used to analyze the relationships between the variables are presented below.

3.1. Classification Results

Two models were used to predict the *department* variable. The first model estimated the department variable using only gender, age, and academic achievement variables. The confusion matrix, obtained by the Random Forest algorithm, is shown in Figure 4.

Predicted

Actual	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**
a	**17**	6	6	2	1	2	2	1
b	5	**46**	7	2	7	8	2	2
c	3	7	**27**	2	2	3	2	0
d	3	4	3	**5**	3	2	4	4
e	2	4	1	2	**10**	4	5	0
f	4	11	4	2	3	**18**	9	11
g	2	3	2	5	4	6	**36**	2
h	1	3	0	7	0	10	3	**19**

a: Mathematics Teaching
b: Classroom Teaching
c: Psyc. Counseling and Guidance
d: Other
e: Social Studies Teaching
f: Turkish Language
g: Comp. Education Teaching
h: Science Teaching

Correctly Classified Instances:
178/383 46.47%

Figure 4. Classification findings related to the first model estimating the department

As seen in Figure 4, the number of Correctly Classified Instances was 178, and the accuracy value was 46.47% out of a total of 383 data points. Students study in eight different departments. Therefore, considering that an eight-variable classifier was used to estimate the department variable, it can be said that the obtained accuracy is an acceptable result.

In the second model, all variables (academic self-efficacy, community feeling, age, gender, and academic achievement) were used as input for the classifier, and the *department* variable was estimated. The Confusion Matrix and classification accuracy obtained are shown in Figure 5.

Predicted

Actual	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**
a	**11**	11	4	1	1	5	2	2
b	4	**38**	19	1	4	8	5	0
c	1	21	**21**	1	0	1	1	0
d	1	5	1	**5**	0	2	9	5
e	1	9	1	1	**6**	6	4	0
f	2	15	2	0	1	**20**	9	13
g	1	2	3	2	2	4	**42**	4
h	0	4	0	1	0	4	2	**32**

a: Mathematics Teaching
b: Classroom Teaching
c: Psyc. Counseling and Guidance
d: Other
e: Social Studies Teaching
f: Turkish Language
g: Comp. Education Teaching
h: Science Teaching

Correctly Classified Instances:
175/383 45.69%

Figure 5. Classification findings related to the second model estimating the department

In Figure 5, the Correctly Classified Instances value is 175, and the accuracy value is 45.69%. This accuracy rate obtained for estimating the quotient variable is also acceptable. It can be concluded that all classification results are acceptable for predicting the department of students.

3.2. Regression Results

To predict the variable of *academic achievement*, we investigated the relationship between academic achievement, demographic data, academic self-efficacy, and community feeling data. The algorithm that yielded the best prediction results was once again Random Forest. Based on the results obtained from the algorithm, the correlation coefficient between academic achievement and the input variables was found to be .55 (Mean absolute error = .34; Root mean squared error = .46; Relative absolute error = 79.92%; Root relative squared error = 83.42%). It can be concluded that these results are acceptable for estimating academic achievement.

3.3. Association Rule Results

The meaningful patterns and rules emerging from the Apriori algorithm benefited from determining the relationships among distance education students' academic self-efficacy, community feelings, academic achievement, and demographic variables, as given in Table 1.

Table 1. Some meaningful patterns and rules emerge with the association rule.

Rules		Conf	Lift
1.	ASE1=*Sufficient* ASE5=*Sufficient* ==> ASE4= *Sufficient*	100%	2.86
2.	ASE3=*Sufficient* ASE4=*Sufficient* CF6=*Medium* ==> ASE2=*Sufficient*	100%	1.85
3.	ASE3=*Sufficient* CF1=*High* CF3=*High* ==> CF4=*High*	97%	2.70
4.	Gender=*Woman* ASE2=*Sufficient* ASE4=*Sufficient* ==>ASE3= *Sufficient*	96%	1.89
5.	ASE6=*Insufficient* CF5=*Low* ==> ASE5= *Insufficient*	96%	1.81
6.	Gender=*Woman* ASE4 ===> ASE3= *Sufficient*	93%	1.82
7.	Difficulty=*No* ASE4=*Sufficient* ASE6=*Sufficient* ==> ASE1=*Sufficient*	93%	1.95
8.	Gender=Woman Difficulty=*No* ASE4=*Sufficient* ==> ASE2=*Sufficient*	92%	1.70

ASE: academic self-efficacy; **CF:** community feeling; **Conf**: confidence value; **Difficulty:** source access difficulty

As seen in Table 1, the rules obtained by the association rule show that EDM can be used to extract hidden, meaningful, and valuable relationships between variables. This is because high-value rules, such as a confidence value of 100% and a lift value of 2.86, were obtained. For example, Rule 8 states that 92% of female students who consider themselves sufficient in the judgment ASE4, and do not have difficulty accessing resources, also consider themselves sufficient in the judgment ASE2. The lift value of Rule 8 (lift = 1.70) indicates the interestingness of this rule, as the lift value is a measure of the importance of a rule. The higher this value is compared to 1, the more valuable the result. This interpretation also applies to the other rules.

4. Conclusions and Suggestions

Traditional statistical methods can lead to the loss or underutilization of important information in critical areas such as education. As a result, alternative analysis methods like data mining have been developed as contemporary approaches to traditional statistical methods. Therefore, the aim of this study is to demonstrate the significance of Educational Data Mining (EDM) through practical application.

In this study, an exemplary application of EDM was conducted to examine the relationships between students' academic achievements, demographic characteristics (age, gender, and department), and social variables such as academic self-efficacy and sense of community. The Apriori algorithm, one of the association rule algorithms in data mining, was employed to analyze these relationships. Additionally, the Random Forest, a classification algorithm, was used to estimate the department and academic achievement variables.

The results of the analysis revealed that the association rule algorithm successfully exposed hidden, interesting, valuable, and meaningful relationships between variables, while the classification algorithms accurately predicted the variables. This outcome demonstrates that data mining methods and algorithms can elucidate and forecast the relationships between various attitudes and behaviors in fields such as educational sciences. In other words, data mining models are applicable to both predictive and descriptive academic research. According to Zaiane [44], data mining is the process by which precise and intriguing patterns can be automatically extracted from large datasets, providing insights into the learning process or student behavior. Injadat et al. [45]

also highlight the potential of data mining methods in analyzing training data. Similarly, in the study conducted by Tekin and Polat [46], the results of the association rule analysis supported the findings of traditional statistical analyses (t-test and ANOVA), and significant patterns that traditional statistical analyses failed to identify were discovered through association rule analysis. Another study [22] revealed that assessments conducted using EDM yielded more comprehensive information about students in higher education.

Data mining methods, in addition to traditional statistical analysis, can enhance research in numerous ways and provide more inferences about the participants [19, 31]. However, data mining can be a complex analysis method that may be challenging for today's educators and administrators to utilize. Therefore, policymakers should encourage the development of software that incorporates data mining, making it accessible for educators and school administrators. Additionally, students should be made aware of EDM. To achieve this, faculty members teaching Scientific Research Methods courses should incorporate both traditional statistical methods and contemporary methods like EDM. Furthermore, a separate course dedicated to EDM should be included in the curriculum at the master's or doctorate levels.

References

[1] Haberal I. Analysis of web log using data mining algorithms. MSc thesis, Baskent University, Ankara, Turkey, 2007.
[2] Chen MS, Han J, Yu, PS. Data mining: an overview from a database perspective. IEEE Trans. Knowl Data Eng 1996; 8(6): 866-883.
[3] Angeli C, Howard SK, Ma J, Yang J, Kirschner PA. Data mining in educational technology classroom research: Can it make a contribution?. Comput Educ 2017; 113: 226-242.
[4] Ateş Y, Karabatak M. (2017). Multiple minimum support values for quantitative association rules. Science and Engineering Journal of Firat University 2017; 29(2): 57-65
[5] Peña-Ayala A. (Ed.) Educational Data Mining: Applications and Trends (Vol. 524). Switzerland: Springer International Publishing, 2014.
[6] Pujari AK. Data Mining Techniques. Hyderabad: Universities Press (India) Private Limited, (2001).
[7] Kaya H, Köymen K. Data mining concept and its application areas. Fırat University Journal of Middle Eastern Studies 2008; 6(2): 159-164.
[8] Thuraisingham B. A primer for understanding and applying data mining. It Professional 2000; 2(1): 28-31.
[9] Velickov S, Solomatine D. (2000). Predictive data mining: practical examples. In 2nd Joint Workshop on Applied AI in Civil Engineering. 2000; 1: 1-17.
[10] Peng Y, Kou G, Shi Y, Chen Z. (2008). A descriptive framework for the field of data mining and knowledge discovery. International Journal of Information Technology & Decision Making 2008; 7(4): 639-682.
[11] Cemaloğlu N, Duykuluoğlu A. Data Mining in Social Sciences. Ankara: PegemA, 2020.
[12] Sevindik T, Kayışlı K, Ünlükahraman O. Data mining in web-based education. Turk J Comput Math Educ 2012; 3(3): 183-193.
[13] Hand DJ, Mannila H, Smyth P. Principles of Data Mining. Massachusetts: MIT Press, 2001.
[14] Asif R, Merceron A, Ali SA, Haider NG. Analyzing undergraduate students' performance using educational data mining. Comput Educ 2017; 113: 177-194.
[15] Chalaris M, Gritzalis S, Maragoudakis M, Sgouropoulou C, Tsolakidis A. Improving quality of educational processes providing new knowledge using data mining techniques. Procedia-Social and Behavioral Sciences 2014; 147: 390-397.
[16] Romero C, Ventura S, Pechenizkiy M, Baker RS. (Eds.). Handbook of Educational Data Mining. USA: CRC Press, 2010.
[17] Murugananthan V, ShivaKumar BL. An adaptive educational data mining technique for mining educational data models in e-learning systems. Indian Journal of Science and Technology 2016; 9(3): 1-5.
[18] Romero C, Ventura S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2010; 40(6): 601-618.
[19] Berland M, Baker RS, Blikstein P. Educational data mining and learning analytics: Applications to constructionist research. Technol, Knowl Learn 2014; 19(1-2): 205-220.
[20] Liu C, Zou D, Chen X, Xie H, Chan WH. A bibliometric review on latent topics and trends of the empirical MOOC literature (2008–2019). Asia Pacific Education Review 2021; 1-20.
[21] Hung HC, Liu IF, Liang CT, Su YS. Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education. Symmetry 2020; 12(2): 213.
[22] Baker RS, Corbett AT. Assessment of robust learning with educational data mining. Research & Practice in Assessment 2014; 9: 38-50.
[23] Şengür D, Karabatak S. Data mining techniques based on students achievements analysis. Turkish Journal of Science & Technology 2018; 13(2): 53-59.
[24] Göksu IB. Learners' evaluation based on data mining in a web-based learning environment. Journal of Computer and Education Research 2015; 3(5): 78-95.
[25] Delavari N, Beikzadeh MR, Phon-Amnuaisuk S. Data mining application in higher learning institutions. Informatics in Education 2008; 7(1): 31-54.
[26] Scheuer O, McLaren BM. Educational data mining. In Encyclopedia of the Sciences of Learning 2012; 1075-1079.
[27] Bilici Z, Özdemir D. Data mining studies in education: Literature review for the years 2014-2020. Journal of Bayburt Education Faculty 2022; 17(33): 342-376.

[28] Baker RS. Data mining for education. International Encyclopedia of Education 2010; 7(3): 112-118.
[29] Charitopoulos A, Rangoussi M, Koulouriotis D. On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010–2018. Int J Artif Intell Educ 2020; 30(3): 371-430.
[30] Mehta MH, Chauhan NC, Gokhale A. Predicting institute graduation rate with genetic algorithm assisted regression for education data mining. ICTACT Journal on Soft Computing 2021;11(2): 2266-2278.
[31] Fernandes CM, Bouthillette F, Raboud JM, Bullock L, Moore CF, Christenson JM, ... Way M. Violence in the emergency department: a survey of health care workers. Cmaj 1999; 161(10): 1245-1248.
[32] Rodrigues MW, Isotani S, Zarate LE. Educational Data Mining: A review of evaluation process in the e-learning. Telematics and Informatics 2018; 35(6): 1701-1717.
[33] Büyüköztürk Ş, Kiliç-Çakmak E, Akgün ÖE, Karadeniz Ş, Demirel F. Scientific Research Methods. Ankara: PegemA, 2011.
[34] Gliner JA, Morgan GA, Leech, NL. Research Methods in Applied Settings: An Integrated Approach to Design and Analysis. Routledge, (2016).
[35] Jerusalem M, Schwarzer R. Fragebogen zur Erfassung von "Selbstwirksamkeit. Skalen zur Befindlichkeit und Persoenlichkeit In R. Schwarzer (Hrsg.). (Forschungsbericht No. 5). Berlin: Freie Universitaet, Institut fuer Psychologie,1981.
[36] Yilmaz M., Gurçay D., Ekici G. Adaptatıon of the Academıc Self-Efficacy Scale to Turkish. H. U. Journal of Education, 2007; 33(33): 253-259.
[37] Rovai AP, Wighting MJ, Lucking R. The Classroom and school community inventory: Development, refinement, and validation of a self-report measure for educational research. The Internet and Higher Education 2004; 7(4): 263-280.
[38] Ilgaz H., Aşkar P. The Development of a Community Feeling Scale toward Online Distance Education Environments. Turkish Journal of Computer and Mathematics Education (TURCOMAT) *2009: 1*(1): 27-34.
[39] Arslan E. (2012). Data mining methods, Hyperlink: [https://emraharslanbm.wordpress.com/2012/07/17/ veri-madenciligi-yontemleri/]. Retrieved on 02 August 2021.
[40] Tanyıldızı E, Karabatak M, Yıldırım G, Özpolat Z. Performance analysis of classification algorithms in wart treatment. Science and Engineering Journal of Firat University 2018; 30(2): 249-256.
[41] Akar Ö, Güngör O. Classification of multispectral images using Random Forest algorithm. Journal of Geodesy and Geoinformation 2013; *1*(2): 139-146.
[42] Alan A, Karabatak M. Evaluation of the factors affecting performance on the data set-classification relationship. Science and Engineering Journal of Firat University 2020; 32(2): 531-540
[43] Uslu, M. Association rule. Hyperlink: [https://www.slideshare.net/uslumetin/ birliktelik-kurallar-kullanlarak-pazar-sepeti-analizi-market-basket-analysis-using-association-rules], 2016. Retrieved on 05 December 2021.
[44] Zaiane OR. Web usage mining for a better web-based learning environment. Proceedings of the 4th IASTED International Conference on Advanced Technology for Education, 2001. https://era.library.ualberta.ca/items/0a182195-ce39-4b5d-a1c1-291ed91a0f36
[45] Injadat M, Moubayed A, Nassif AB, Shami A. Systematic ensemble model selection approach for educational data mining. Knowledge-Based Syst, 2020, 105992.
[46] Tekin A, Polat E. Evaluation of teacher candidates' techno-pedagogical education competencies with the rule of association. In: Nabiyev, V, Erümit AK, editors. Artificial Intelligence in Education From Theory to Practice. Ankara: PegemA, 2020.

APPENDIX -1

Academic Self-Efficacy Scale

Code	**Original Form**	**Adapted Form**
ASE1	I am always in a position to accomplish the things that need to be done in my university education.	Üniversite öğrenimimde her zaman yapılması gereken işleri başarabilecek durumdayım
ASE2	I always achieve high success when I am adequately prepared for the exam	Yeterince hazırlandığım zaman sınavlarda daima yüksek başarı elde ederim
ASE3	I know very well what I need to do to get good grades	İyi not almak için ne yapmam gerektiğini çok iyi biliyorum
ASE4	Even if a written exam is very difficult, I know that I will pass it	Bir yazılı sınav çok zor olsa bile, onu başaracağımı biliyorum
ASE5	I cannot think of failing any exam	Başarısız olacağım herhangi bir sınav düşünemiyorum
ASE6	I have a relaxed attitude in exam environments because I trust my intelligence	Sınav ortamlarında rahat bir tavır sergilerim, çünkü zekama güveniyorum
ASE7	I usually do not know how to deal with the subjects I need to learn while preparing for exams	Sınavlara hazırlanırken öğrenmem gereken konularla nasıl başa çıkmam gerektiğini genellikle bilemem.(-)

Community Feelings Scale

Code	**Original Form**	**Adapted Form**
CF1	I feel that students in this program care about each other	Bu programdaki öğrencilerin birbirlerini önemsediklerini hissederim.
CF2	I trust others who take this course	Bu dersi alan diğer kişilere güvenirim.
CF3	I think this program meets my educational needs	Bu programın eğitim ihtiyaçlarımı karşıladığını düşünüyorum.
CF4	I think this program gives me a lot of opportunities to learn	Bu programın öğrenmem için bana çokça fırsat verdiğini düşünüyorum.
CF5	I regularly talk to those in this program about my personal issues	Bu programdakilerle kişisel konularım hakkında düzenli olarak konuşurum.
CF6	I share educational values with others in this program	Bu programdaki diğer kişilerle eğitsel değerleri paylaşırım.