

## Investigation of the relationships of the students' academic level and gender with Covid-19 based anxiety and protective behaviors: A data mining approach

Dönüş ŞENGÜR\*

Department of Educational Sciences, Faculty of Education, Firat University, Elazığ, Turkey

\*dsengur@firat.edu.tr

(Geliş/Received: 07/06/2020;

Kabul/Accepted: 13/08/2020)

**Abstract:** COVID-19, a new virus, has been caused an outbreak in all around the world. It has affected almost all parts of our lives such as working habits. The life of human beings has come to a halt. People have started to work in their homes. Besides, education activities from preschool to the colleges have canceled and distance courses took place instead of face to face education. This situation has caused fear and anxiety. Especially for students, the anxiety level about education has increased during the pandemic. In this study, the relationship of the students' academic level and gender with Covid-19 based anxiety and protective behaviors is investigated based on the data mining approach. To this end, an association rule-based classification (ARC) method is employed. Moreover, other classification approaches namely decision trees (DT), support vector machines (SVM) and k-nearest neighbor (k-NN) are also used. The ARC is used to detect the association rules between items of the dataset and obtained rules are used to construct a classifier. To detect the relationships between the students' academic level and gender with COVID-19 based anxiety and protective behaviors, a dataset, which was constructed from 215 university students by using an online self-administered questionnaire, is considered in experimental studies. The dataset covers three instruments namely Anxiety Scale (AS) with 10 items, Protective Behaviors Scale (PBS) with 14 items, and Related Knowledge Scale (RKS) with 12 items, respectively. The experimental results show that the proposed data mining approaches produce satisfactory results in determining the relationship between the students' academic level and gender with Covid-19 based anxiety and protective behaviors.

**Key words:** COVID-19, students' anxiety level, data mining, prediction, rule mining.

### Öğrencilerin akademik düzeyi ve cinsiyetinin Covid-19 temelli kaygı ve koruyucu davranışlarla ilişkisinin incelenmesi: Bir veri madenciliği yaklaşımı

**Öz:** Yeni bir virüs olan COVID-19, tüm dünyada salgına neden oldu. Çalışma alışkanlıkları gibi hayatımızın hemen hemen her alanını etkiledi. İnsanın hayatı durma noktasına geldi. İnsanlar evlerinde çalışmaya başladılar. Ayrıca okul öncesi eğitimden kolejlere kadar olan eğitim faaliyetleri iptal edilmiş ve yüz yüze eğitim yerine uzaktan eğitimler verilmiştir. Bu durum korkuya ve endişeye neden oldu. Pandemi sırasında özellikle öğrenciler için eğitimle ilgili kaygı düzeyi artmıştır. Bu çalışmada öğrencilerin akademik düzeyi ve cinsiyetinin Covid-19 temelli kaygı ve koruyucu davranışlarla ilişkisi veri madenciliği yaklaşımına dayalı olarak incelenmiştir. Bu amaçla, bir ilişki kuralı tabanlı sınıflandırma (KTS) yöntemi kullanılır. Ayrıca, karar ağaçları (KA), destek vektör makineleri (DVM) ve k-en yakın komşu (k-NN) gibi diğer sınıflandırma yaklaşımları da kullanılmaktadır. KTS, veri kümesinin öğeleri arasındaki ilişki kurallarını tespit etmek için kullanılır ve elde edilen kurallar bir sınıflandırıcı oluşturmak için kullanılır. Öğrencilerin akademik düzeyi ve cinsiyeti ile COVID-19 temelli kaygı ve koruyucu davranışlar arasındaki ilişkiyi saptamak için, 215 üniversite öğrencisinden çevrimiçi kendi kendine uygulanan bir anket kullanılarak oluşturulmuş bir veri seti deneysel çalışmalarda ele alınmıştır. Veri seti, sırasıyla 10 maddelik Kaygı Ölçeği (KÖ), 14 maddelik Koruyucu Davranışlar Ölçeği (KDÖ) ve 12 maddelik İlgili Bilgi Ölçeği (İBÖ) olmak üzere üç aracı kapsamaktadır. Deneysel sonuçlar, önerilen veri madenciliği yaklaşımlarının, öğrencilerin akademik düzeyi ve cinsiyeti ile Covid-19 temelli kaygı ve koruyucu davranışlar arasındaki ilişkiyi belirlemede tatmin edici sonuçlar ürettiğini göstermektedir.

**Anahtar kelimeler:** KOVID-19, öğrencilerin kaygı düzeyleri, veri madenciliği, tahmin, kural madenciliği.

#### 1. Introduction

COVID-19, which caused a pandemic worldwide, has been first appeared in Wuhan city of China in the last days of 2019 [1]. Although the disease initially looked like the ordinary flu, over time it became clear how dangerous the disease was. The disease has had many negative effects on our lives. Due to the disease, some concepts such as working from home, education with distance courses and curfew became to be a part of our lives. Students, who are an enormous group, affected by the COVID-19 pandemic. Students have started to get their courses by distance courses and have had to stay at home due to the curfew. This situation brings anxiety and

\* Corresponding author: dsengur@firat.edu.tr. ORCID Number of author: 0000-0002-8786-6557

mental tiredness to the students. During the pandemic term, researchers have investigated how to measure the psychology situation of the students based on some instruments [2]. Riad et al. [2] developed instruments to evaluate COVID-19 based anxiety level, protective behaviors, and knowledge towards the COVID-19. A self-administered questionnaire was developed and used for data collection from 215 university students who were from Europe and China via social networks, respectively. The authors used item-total analysis and confirmatory factor analysis models for evaluating the collected dataset. The results showed that the developed instruments were promising for measuring the psychological effects of COVID-19. Trung et al. [3] investigated the learning habits of the Vietnamese students during the COVID-19 pandemic. The authors spread an online questionnaire on a network of educational communities over social networks to construct a dataset. Cao et al. [4] evaluated the mental situation of medical school students in China during the COVID-19 pandemic. A 7-item questionnaire, which examines the anxiety disorder levels of the participant, was used to collect the dataset. 7143 students were recruited in the dataset. Several statistical analyses were carried out on the collected dataset and some important findings were reported. Odriozola-González et al. [5] evaluated the mental impact of COVID-19 on university students during the pandemic lockdown in Spain. Anxiety, depression and stress levels of the students were measured and collected in a dataset by using an online questionnaire. 2530 students from the University of Valladolid were recruited in the study. The statistical analysis over-collected dataset revealed that students were suffered an important mental impact during the pandemic. Bitan et al. [6] investigated the degree of the fear factor of COVID-19 in Israel. Besides fear, anxiety, depression and stress level of the participants were saved into a dataset. 639 participants were used in the construction of the dataset and the two-factor structure model was used in the evaluation of the collected dataset. Authors reported that gender and chronic illness were two important factors associated with the fear of COVID-19. Zhang et al. [7] evaluated the adverse impact of the COVID-19 pandemic on Chinese students' anxiety and depression levels. An online questionnaire, which measured the demographics, physical activity, negative emotions, sleep quality, and aggressiveness level of the participants, were used on 66 students to construct the dataset. The authors used a mixed-effect model to evaluate the associations between variables. The authors also investigated the mediating effect of sleep quality. Kaparounaki et al. [8] explored students' psychological state during the pandemic lockdown. An online questionnaire was used to collect the dataset over 1000 university students. The depression, anxiety level and suicidal thoughts of the students were recorded by using the questionnaire. The results showed that there have been increases in anxiety, depression and suicidal thoughts, respectively.

As mentioned earlier, COVID-19 has been affected the whole world in almost all parts such as economy, lifestyle, education, travel etc. Education, which is an important part of our lives, has affected deeply. Especially, students' anxiety level, protective behaviors and knowledge scales have been changed due to the pandemic. In this work, it was aimed to shown the relationship between students' anxiety level, protective behaviors and knowledge scales and academic level and gender during the pandemic. To this end, various datamining techniques have been used for prediction purposes. As the recent literature is reviewed, it is seen that the researchers generally applied online questionnaires and collected datasets for measuring the students' mental health during the COVID-19 pandemic. In this study, data mining and machine learning techniques are used to determine the relationship between the students' academic level and gender with COVID-19 based anxiety and protective behaviors. The dataset, which was collected in [2], is considered in this study. 215 university students were recruited during the dataset collection where an online self-administered questionnaire was used. The questionnaire covered three instruments namely Anxiety Scale (AS) with 10 items, Protective Behaviors Scale (PBS) with 14 items, and Related Knowledge Scale (RKS) with 12 items, respectively. ARC method, which is generally used to detect the association rules between items of the dataset, is employed for gender and academic level prediction. Besides, DT, SVM and k-NN methods are further used for prediction and comparison purposes. The accuracy is used as the measure of validation of the proposed work. The obtained results reveal that data mining and machine learning are a promising tool for such applications.

The remainder of this paper is as follows. The next section introduces the background theories. The apriori algorithm and ARC theories are briefly introduced in Section 2. Section 3 describes the dataset, experimental works and results. The paper is concluded in Section 4.

## 2. Background Theories

This section briefly introduced the theories of the data mining approaches namely the apriori algorithm and ARC method [9, 10]. Readers may refer to [11, 12] for detailed information about the mentioned methods.

## 2.1. Apriori Algorithm

Apriori algorithm is known as the basic and ancient method in association rule mining [9]. Many new approaches have been developed based on the concept of the Apriori algorithm [13-15]. Apriori algorithm mines the frequent itemsets in a given set of transactions. Apriori algorithm produces the rules based on the support and confidence values. While the support is defined as the ratio of the total number of records of transactions that include all items in the antecedent and consequent parts of the rule, the confidence is defined as the percentage of the number of transactions that include all items in the consequent, as well as the antecedent, to the number of transactions that include all items. Lift is another metric that is used to evaluate the significance and reliability of a rule. The rules are generally in the form of  $A \rightarrow B$  where A shows the antecedent part of the rule and B indicates the consequent part of the rule, respectively. Equations 1, 2, 3 and 4 show the rule, support, confidence and lift measures, respectively.

$$\text{Rule: } A \rightarrow B \quad (1)$$

$$\text{Support} = \frac{\text{freq}(A, B)}{N} \quad (2)$$

$$\text{Confidence} = \frac{\text{freq}(A, B)}{\text{freq}(A)} \quad (3)$$

$$\text{Lift} = \frac{\text{Support}}{\text{Support}(A) \cdot \text{Support}(B)} \quad (4)$$

where ‘freq’ shows the frequency.

## 2.2. Associative rules-based classification

ARC is a supervised rule-based classifier, which uses the rules that are obtained by the apriori algorithm [16]. Initially, the frequent items that meet the minimum confidence threshold are saved for rule-based classifier construction. The saved rules, whose consequent parts are restricted to the class label, are further selected for classification purposes. A heuristic approach is employed for ranking the selected rules according to their confidence values and the rules that cover the training samples are selected.

The ARC algorithm initially selects the best rule and then eliminates all the covered examples. If at least one example meets the rule conditions, this rule will be added to the final rules. This procedure is repeated until there are no more rules to select or no other samples within the scope. The algorithm then stops and returns the classifier in the form of the if-then-else rule list.

## 3. Experimental studies and results

### 3.1. Dataset

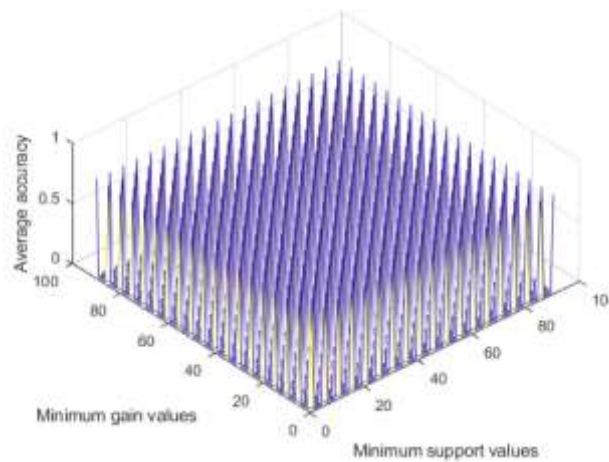
The students' anxiety level and protective behaviors dataset were constructed based on 215 university students from 17 counties [2]. An online, functionality and user-friendly questionnaire, which was spread between March 25th-27th 2020 via social networks, contained demographic information and three instruments namely anxiety scale with 10 items, protective behaviors scale with 14 items, and knowledge scale with 12 items. For anxiety and protective behavior scales, a 5-point Likert scales with 10 and 14 items were used, respectively. For anxiety scale "1" referred to "Totally disagree" and "5" referred to the "Totally agree". Similarly, for protective behaviors scale, "1" referred to "Not at all like me" and "5" referred to "Just like me". For the knowledge scale, a multi-choice approach was adopted. Each item has four available options where one of them is the right option. More detailed information can be seen in [2]. The distributions of the recordings among the classes such as academic level and gender are given Table 1.

**Table 1.** The distribution of the recordings among the classes

Academic level	Bachelor	Masters	Doctoral
	152	49	14
Gender	Male		Female
	173		42

### 3.2 Results and Discussions

All coding was carried out with MATLAB on a computer equipped with NVIDIA Quadro M4000 GPU and Intel(R) Xeon(R) CPU E5-1650 @3.60 GHz 64 GB memory. Five-fold cross-validation technique and average accuracy score were used in experiments for evaluation of the proposed method. The accuracy is defined as the ratio of the number of correct classifications to the number of all samples in the dataset. In the experiments, both academic level and gender were used as output. While gender output has two labels, the academic level has three labels such as “bachelor”, “masters” and “doctoral”, respectively. Besides, anxiety, protective behaviors and knowledge scales were both used independently and concatenated forms. For ARC, the selection of the minimum support and gain values were carried out with a grid search algorithm. The minimum gain value was used for selecting the best rules. The support values were in the range of  $10^{-2}$  and 1 and the gain values were in the range of  $10^{-3}$  and 2, respectively. Moreover, the number of the best rules was fixed to five. This value was found heuristically during the experimental works. The initial experiments were carried out on an anxiety scale. Figure 1 shows the grid search results for anxiety scale vs gender classification. While the x-axis shows the minimum support values, the y-axis shows the minimum gain values. The z-axis indicates the accuracy scores.



**Figure 1** Grid search on an anxiety scale

Table 2 also shows the average accuracy scores for anxiety scale based academic level and gender classification. As seen in Table 2, by using anxiety scale, the academic level and the gender of the participants were classified with 73.95% and 79.93% average accuracy scores, respectively.

**Table 2** Average accuracy scores for Anxiety scale

Anxiety Scale	Average Accuracy
Academic level	73.95%

Gender	79.93%
--------	--------

Table 3 gives the average prediction scores for the scale of protective behavior. As seen, 70.70% and 80.40% average accuracy scores were obtained for academic level and gender, respectively. Gender classification was carried out better than the academic level classification.

**Table 3** Average accuracy scores for protective behaviors scale

Anxiety Scale	Average Accuracy
Academic level	70.70%
Gender	80.40%

Table 4 was constructed for giving the classification achievements for the knowledge scale. While the academic level was predicted with a 66.50% average accuracy score, 77.21% score was produced for gender class.

**Table 4** Average accuracy scores for knowledge scale

Anxiety Scale	Average Accuracy
Academic level	66.50%
Gender	77.21%

From Tables 2, 3 and 4, it was seen that for single scales, the scale of the protective behavior produced the high average accuracy score for gender class and the academic level was predicted better with the anxiety scale. It is worth mentioning that the knowledge scale generally produced low average accuracy score when compared with the anxiety and protective behaviors scales, respectively.

**Table 5** Average accuracy scores for anxiety + protective behaviors scales

Anxiety Scale	Average Accuracy
Academic level	74.16%
Gender	80.47%

In Table 5, the results that were obtained via the concatenation of the anxiety and protective behavior scales were given. One important observation was that the concatenation of the scales improved the classification accuracy scores. 74.16% and 80.47% average accuracy scores were obtained for academic level and gender classes, respectively.

**Table 6** Average accuracy scores for anxiety + knowledge scales

Anxiety Scale	Average Accuracy
Academic level	74.16%
Gender	79.07%

The average accuracy scores for concatenated anxiety and knowledge scales were given in Table 6. As compared in Tables 5 and 6, both concatenated anxiety and protective behaviors and anxiety and knowledge scales produced the identical average accuracy scores for academic level class. Besides, a 79.07% average accuracy score was produced for the gender class as seen in Table 6.

**Table 7** Average accuracy scores for protective behaviors + knowledge scales

Anxiety Scale	Average Accuracy
Academic level	75.77%
Gender	80.93%

The achievements for concatenated protective behaviors + knowledge scales were given in Table 7. As seen in Table 7, a 75.77% average accuracy score was obtained for academic level class and an 80.93% average accuracy score was produced for gender class.

**Table 8** Average accuracy scores for anxiety + protective behaviors + knowledge scales

<b>Anxiety Scale</b>	<b>Average Accuracy</b>
Academic level	79.43%
Gender	82.33%

Finally, the achievements for concatenation of all scales were given in Table 8. A 79.43% average accuracy score was produced for academic level and an 82.33% average accuracy score was obtained for gender class.

Comparison of the obtained results for concatenation of the anxiety, protective behaviors and knowledge scales with some of the existing machine learning techniques were tabulated in Table 9. MATLAB Classification Learner Tool (MCLT) was used for comparison purposes. The default setting parameters of compared methods in MCLT was used. DT is a flow chart, such as a tree structure, where each inner node represents a test on an attribute, each branch represents a result of the test, and each leaf node (terminal node) contains a class tag [17]. Given a set of training examples, each marked as one or the other of two categories, an SVM training algorithm makes new examples to one category or another making it an unlikely binary linear classifier [18]. The output in the k-NN classification is a class member. An object is classified by the plural vote of its neighbors; the object is assigned to the class most common among its closest neighbors. If  $k = 1$ , the object is assigned to the class of the nearest neighbor [19].

**Table 9** Performance comparison of average accuracy scores for anxiety + protective behaviors + knowledge scales

	<b>Academic level</b>	<b>Gender</b>
<b>DT</b>	61.4%	71.2%
<b>SVM</b>	74.9%	81.4%
<b>k-NN</b>	72.6%	80.5%
<b>ARC</b>	79.4%	82.3%

From Table 9, it is observed that the ARC produced the highest average accuracy scores for both academic level and gender. The second-best average accuracy scores were produced by the SVM technique. SVM produced 74.9% and 81.4% average accuracy scores for academic level and gender classes, respectively. k-NN yielded the third-best accuracy scores where 72.6% and 80.5% average accuracy scores were obtained for academic level and gender. Finally, the worst results, where 61.4% and 71.2% average accuracy scores produced for academic level and gender classes, were obtained by the DT method.

#### 4. Conclusions

In this paper, the prediction of the students' academic level and gender with COVID-19 based anxiety and protective behaviors was investigated based on data mining and machine learning approach. A rule-based classifier namely ARC method was employed for this purpose. Besides, DT, SVM and k-NN methods were also used. The ARC was used to detect the association rules between items of the dataset and obtained rules are used to construct a classifier. To detect the relationships between the students' academic level and gender with COVID-19 based anxiety and protective behaviors, a dataset, which was constructed from 215 university students by using an online self-administered questionnaire, is considered in experimental studies. The following conclusions are extracted from the study.

1-) From the experimental analysis, it was seen that data mining and machine learning techniques have potential in the use of finding relationships between COVID-19 based anxiety, protective behaviors and knowledge scales and the academic level and gender of the students.

2-) Prediction of the gender was achieved better than the prediction of the academic level. This might be based on the number of class labels as the academic level has three class labels.

3-) Using multi scales has increased the prediction performance greatly.

In the future works, we are planning to construct a dataset for Turkish students and examine the effect of the COVID-19 on Turkish students' anxiety level and protective behaviors.

### References

- [1] World Health Organization. IHR Emergency Committee on Novel Coronavirus (2019-nCoV) (2020).
- [2] Riad A, Huang Y, Zheng L, & Elavsky S. COVID-19 Induced Anxiety and Protective Behaviors During COVID-19 Outbreak: Scale Development and Validation. medRxiv 2020.
- [3] Trung T, Hoang AD, Nguyen TT, Dinh VH, Nguyen YC, & Pham HH. Dataset of Vietnamese student's learning habits during COVID-19. Data in Brief 2020; 105682.
- [4] Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, & Zheng J. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry research 2020; 112934.
- [5] Odriozola-González P, Planchuelo-Gómez Á, Irurtia MJ, & de Luis-García R. Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. Psychiatry Research 2020; 113108.
- [6] Bitan DT, Grossman-Giron A, Bloch Y, Mayer Y, Shiffman N, & Mendlovic S. Fear of COVID-19 scale: Psychometric characteristics, reliability and validity in the Israeli population. Psychiatry Research 2020; 113100.
- [7] Zhang Y, Zhang H, Ma X, & Di Q. Mental Health Problems during the COVID-19 Pandemics and the Mitigation Effects of Exercise: A Longitudinal Study of College Students in China. International Journal of Environmental Research and Public Health 2020; 17(10), 3722.
- [8] Kaparounaki CK, Patsali ME, Mousa DPV, Papadopoulou EV, Papadopoulou KK, & Fountoulakis KN. University students' mental health amidst the COVID-19 quarantine in Greece. Psychiatry Research 2020; 113111.
- [9] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on management of data, Washington DC, 25-28 May 1993. New York: ACM, 207–216
- [10] Cano A, Zafra A, Ventura S. An interpretable classification rule mining algorithm. Information Sciences 2013; 240: 1–20.
- [11] Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo AI. Fast discovery of association rules. Advances in Knowledge Discovery and Data Mining 1996; 12: 307–328.
- [12] Hasanpour H, Meibodi RG, & Navi K. Improving rule-based classification using Harmony Search. PeerJ Computer Science 2019; 5: e188.
- [13] Scheffer T. Finding association rules that trade support optimally against confidence. Principles of Data Mining and Knowledge Discovery 2001; 424–435.
- [14] Nahar J, Imam T, Tickle KS, Chen Y-PP. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications 2013; 40: 1086–1093.
- [15] Nahar J, Tickle KS, Ali AS, Chen Y-PP. Significant cancer prevention factor extraction: an association rule discovery approach. Journal of Medical Systems 2011; 35: 353–367.
- [16] Ma BLWHY, Liu B. Integrating classification and association rule mining. Proceedings of the fourth international conference on knowledge discovery and data mining, 1998.
- [17] Şengür D, Tekin A. Prediction of Student's Grade Point Average by Using the Data Mining Methods, International Journal Of Informatics Technologies 2013; 6 (3) 7-16
- [18] Vapnik VN (1995) The Nature of Statistical Learning Theory; Springer, New York, USA.
- [19] Şengür D, Turhan M. Prediction of the action identification levels of teachers based on organizational commitment and job satisfaction by using k-nearest neighbors method. Firat Univ Turkish J Sci Technol. 2018;13(2):61–8.