EDUCATIONAL DATA MINING STUDIES IN TURKEY: A SYSTEMATIC REVIEW

Dr. Seyhmus AYDOGDU
ORCID: 0000-0002-9075-8055
Department of Computer Education and Instructional Technologies
Nevsehir Haci Bektas Veli University
Nevsehir, TURKEY

ABSTRACT

The purpose of this research is a comprehensive review of studies towards educational data mining (EDM) in Turkey. For the purpose of this study, graduate theses and articles conducted in Turkey were examined in detail. As a result of the literature review, 48 studies were analyzed in the context of the data mining purpose, the technique used in the study, the purpose of the study, the sample, the data sources and the analysis tool used in the study. In addition, the input variables used in the studies for estimating student achievement in EDM were examined. As a result of the research, it was found that EDM studies were mostly aimed at prediction from data mining tasks. One of the results of this study is that artificial neural networks are the most commonly used technique in EDM studies. When the distributions of EDM studies according to their objectives are examined, it is seen that studies are predominantly aimed at predicting student achievement. It is seen that university students are preferred as the sample in EDM studies, achievement scores are used as data source and SPSS application is used more as an analysis tool. In the last part of the research, recommendations were made based on the results to the researchers.

Keywords: Educational data mining, data mining, e-learning, review, classification, prediction

INTRODUCTION

Learning management systems and massive open online courses are now being used by various institutions to meet student needs (You, 2016). Educational data mining (EDM) is an interdisciplinary field in which data mining methods are applied to discover data in these systems and make future predictions (Romero & Ventura, 2013). EDM is a new and growing field of research in which meaningful information is extracted by applying data mining techniques on raw data in education systems (Bousbia & Belamri, 2014). EDM is used to provide information to managers and trainings on student interactions in a learning environment, to provide information on how to develop course content in the environment, to advise students on appropriate content, to predict students' achievements, and to be used for student modeling (Romero, Ventura, Pechenizkiy, & Baker, 2010). In order to achieve these objectives, data mining and machine learning methods are used in an integrated way (Baker, 2010; Romero & Ventura, 2010).

EDM can be used for (Berry & Linoff, 2004) classification (Baradwaj & Pal, 2012; Kaur, Singh, & Josan, 2015; Kamisli Ozturk, Erzurum Cicek & Ergul, 2017), prediction (Bucos & Dragulescu, 2018; Francis & Babu, 2019), clustering (Abu Tair & El-Halees, 2012; Dutt, Aghabozrgi, Ismail, & Mahroeian, 2015) and creating association rules (Avlijas, Heleta, & Avlijas, 2016; Man, Abu Bakar, & Sabri, 2018). EDM studies may address one or more of these objectives (Pena-Ayala, 2014).

a. Classification: In the classification, the data is grouped according to the determined features. For example some tasks are examined within the scope of classification such as which content is displayed according to learning styles and which characteristics of students are successful in the course.

b. Prediction: In the prediction, the data is separated into training and test data. Then, a model is formed by using training data and the accuracy rate of the model is examined with the test data. In the prediction, we can use categorical or continuos variables. For example, in an online learning environment, estimating final scores based on students' browsing behavior is a prediction process.

c. Clustering: Clustering is the process of dividing the data in the data set into homogeneous groups as in the classification. There are no predefined groups in the clustering process as in the classification process. For example, student profile creation based on student navigation in the e-learning environment is within the clustering process.

d. Association rules: Association rules, as the name suggests, is the process of generating rules from data. These rules take into account the frequency of coexistence of variables.

In the EDM studies, data mining techniques are applied for the specified purpose. In the literature, according to the purpose of the EDM; decision trees (Lin, Yeh, Hung, & Chang, 2013), bayes classifiers (Bhardwaj & Pal, 2012), regression (Bahadir, 2016), artificial neural networks (Yang & Li, 2018) and support vector machines (Al-Shehri et al., 2017) techniques are widely used.

There are different national and international screening studies related to educational data mining. In the research conducted by Romero and Ventura (2007) which reviewed studies between 1995 and 2005, it was emphasized that educational data mining is a new and developing research area. Three years later, the study by Romero and Ventura (2010) again showed an increase in studies on educational data mining. It is seen that traditional learning environments, e-learning environments, learning management systems, intelligent tutoring systems, adaptive educational systems, tests, questionnaires, texts and contents are used as data sources. In the classification made according to the purpose of educational data mining the studies were grouped as "analysis and visualization of data", "providing feedback for supporting instructors", "recommendations for students", "predicting student's performance", "student modeling", "detecting undesirable student behaviors", "grouping students", "social network analysis", "developing concept maps", "constructing courseware" and "planning and scheduling". In the study, the development of easy-to-use educational data mining environments for non-expert users, the development of data mining tools that work in integration with e-learning environments, storing and publishing models and data in a reusable manner to certain standards and the regulation of traditional data mining algorithms in the educational context were mentioned as future works.

Mohamad and Tasir (2013) examined 9 researches on data mining between 2004-2012. As a result of the study, it is understood that in addition to the use of available learning management systems (Moodle, Blackboard, etc.) in the researches, there are also systems developed within the scope of the research. In the study, it is suggested that social environments such as blog and facebook should be used for collaborative learning and the navigation data should be analyzed by integrating tools such as Google Analytics. Pena-Ayala (2014) reviewed 240 studies towards EDM between 2010-2013 1Q. In this research EDM works are grouped as EDM approaches and EDM tools. EDM approaches consists of "student modeling", "student behavior modeling", "student performance modeling", "assessment", "student support and feedback" and "curriculum, domain knowledge, sequencing and teachers support". EDM tools consist of "extraction, learning support and feature engineering, "visualization" and "analysis support". As a result of the study, it is emphasized that EDM is well-grounded and a new research area, which requires an interdisciplinary perspective.

Manjarres, Sandoval, and Suarez (2018) examined the researches between 2003 and 2015 according to the algorithms used in the EDM and the objectives of the EDM. As a result of the review, decision trees, classification, clustering, bayesian networks, artificial neural networks and association rules techniques are widely used in EDM. Tekin and Oztekin (2018) examined the studies on educational data mining between 2006-2016. In the study, Google Scholar, ERIC (Education Resources Information Center), Elsevier, Proquest, Ulakbim, Ebscohost and Springer databases were searched. Descriptive analysis of 140 articles examined in terms of different variables are presented. When the results of the research are examined, it is seen that the most researched topic is student performance, the participants are generally chosen from university students and web environments are preferred as data collection tools. Aldowah, Al-Samarraie, and Fauzy (2019) examined 402 EDM studies in higher education between 2000 and 2017. In the study, it was stated that student centered strategies could be developed with EDM.

Although there has been some review studies about EDM, it is seen that the limited number of studies conducted on this area in Turkey. In this study, we tried to determine the current situation of educational data mining in Turkey through examine master theses, doctoral dissertations and papers in depth. The

purpose of this research is a comprehensive review of studies towards educational data mining in Turkey. For this purpose, Turkey addressed graduate theses and articles have been examined. For the purpose of the study, the following research questions were answered:

a. How are EDM studies distributed by years?
b. How are EDM studies distributed according to data mining tasks?
c. Which techniques are used in EDM studies?
d. What are the topics in EDM studies?
e. Which sample characteristics chosen in EDM studies?
f. What are the data sources in EDM studies?
g. Which analysis tools are used in EDM studies?
h. Which variables are used as input variables in the researches that predict success in EDM?

This study focuses on studies related to educational data mining in Turkey. The structure of this paper is organized as follows: in the following section the methodology of the study is described. In third section, results of review are presented. Finally, the results were discussed within the framework of the relevant literature and suggestions were made.

METHOD

Search Strategy

Council of Higher Education (CoHE) Thesis Center, Web of Science database and TR Index search tool were used in the review of the studies (Table 1). No date range is specified during the searching process. The review process was terminated in July 2019.

Table 1. Databases used in review process

Database name	Address
CoHE Thesis Center	https://tez.yok.gov.tr/UlusalTezMerkezi/
Web of Science	http://webofknowledge.com
TR Index Search Tool	https://trdizin.gov.tr/

In literature search process "data mining", "machine learning", "neural networks", "decision trees", "bayes" and "support vector machine" keywords were used. The searching using the CoHE Thesis Center was limited to the subject area of Education and Training. The initial search returned 67 studies. In Web of Science tool, "OR" boolean operator was used between keywords and Turkey addressed educational researches was searched. 62 articles were initially accessed in Web of Science. Using the TR Index Search tool, the keywords were searched respectively and the results were filtered by education subject area. Then, 25 studies were reached by eliminating repetitive studies.

Inclusion/Exclusion Criteria

In this study, the "Publication Classification Form" which was developed by the researcher was used to record the searched publications. The studies that were searched were determined in accordance with the inclusion/exclusion criteria and these studies were examined in detail. Studies which do not include any of the data mining tasks (classification, estimation, association rules, clustering), which are literature surveys, which are not educational research and which are not available in CoHE Thesis Center were excluded from the research. In addition, test equating, imputation of missing data, and comparison of two different methods in data mining were not included in this study because they were considered a different subject area.

Data Analysis

After examining the studies, the number of studies that meet the criteria are presented in Table 2.

Table 2. Distribution of accessed studies according to sources

Source	Number of Studies	Appropriate number of studies
CoHE Thesis Center	67	23
Master thesis	38	16
Doctoral dissertation	29	7
Web of Science	62	22
TR Index Search Tool	25	7*
Total	154	52**

*Note. * In addition to these studies, three studies were found in the Web of Science search results.,*
* *Four of these studies were not included in the results because they were produced from the thesis work of the authors.*

As a result, 23 thesis studies, 16 of which were master's and 7 of which were doctoral theses, were included in the analysis. (Table 2). After eliminating the repeated studies from Web of Science and TR Index Search Tool, 29 articles were found to meet inclusion criterias. The list of works that meet the criteria is presented in Table 3.

Table 3. EDM studies included in the research

No	Reference	Tool	Type/Journal Name
1	Aydin (2007)	CoHE Thesis Center	Doctoral dissertation
2	Somyurek (2008)	CoHE Thesis Center	Doctoral dissertation
3	Ucgun (2009)	CoHE Thesis Center	Master thesis
4	Y. Aydogdu (2011)	CoHE Thesis Center	Master thesis
5	Birtil (2011)	CoHE Thesis Center	Master thesis
6	Yelegin (2012)	CoHE Thesis Center	Master thesis
7	Yucel (2012)	CoHE Thesis Center	Master thesis
8	Bahadir (2013)	CoHE Thesis Center	Doctoral dissertation
9	Coskun (2013)	CoHE Thesis Center	Master thesis
10	Hark (2013)	CoHE Thesis Center	Master thesis
11	Saygili (2013)	CoHE Thesis Center	Master thesis
12	Sengur (2013)	CoHE Thesis Center	Master thesis
13	Akcapinar (2014)	CoHE Thesis Center	Doctoral dissertation
14	Aksoy (2014)	CoHE Thesis Center	Master thesis
15	Yildiz (2014)	CoHE Thesis Center	Doctoral dissertation
16	Ozbay (2015)	CoHE Thesis Center	Master thesis
17	Uysal (2015)	CoHE Thesis Center	Master thesis
18	Cebi (2016)	CoHE Thesis Center	Doctoral dissertation
19	Yildiz Aybek (2016)	CoHE Thesis Center	Master thesis
20	Barngrover (2017)	CoHE Thesis Center	Master thesis
21	Sahin (2018)	CoHE Thesis Center	Doctoral dissertation
22	Yagci (2018)	CoHE Thesis Center	Master thesis
23	Yorganci (2018)	CoHE Thesis Center	Master thesis
24	Dogan and Camurcu (2010)	Web of Science	Computer Applications in Engineering Education
25	Guruler, Istanbullu, and Karahasan (2010)	Web of Science	Computers & Education
26	Kentli and Sahin (2011)	Web of Science	Energy Education Science and Technology Part B-Social and Educational Studies
27	Aydogdu and Tanrikulu (2013)*	Web of Science	Education and Science

28	Turhan, Kurt, and Engin (2013)	Web of Science	Education and Science
29	Akcapinar (2015)	Web of Science	Computers & Education
30	Avsar and Yalcin (2015)	Web of Science	Education and Science
31	Baran and Kilic (2015)	Web of Science	Educational Technology & Society
32	Demir (2015)	Web of Science	Educational Sciences-Theory & Practice
33	Kayri (2015)	Web of Science	Educational Sciences-Theory & Practice
34	Sohsah, Unal, and Guzey (2015)	Web of Science	British Journal of Educational Technology
35	Bahadir (2016)	Web of Science	Educational Sciences-Theory & Practice
36	Idil, Narli, and Aksoy (2016)	Web of Science	International Journal of Education in Mathematics Science and Technology
37	Kose and Arslan (2016)	Web of Science	International Journal of Engineering Education
38	Onan, Bal, and Bayam (2016)	Web of Science	Croatian Journal of Education
39	Ayvaz, Guruler, and Devrim (2017)	Web of Science	Information Technologies and Learning Tools
40	Kose and Arslan (2017)	Web of Science	Computer Applications in Engineering Education
41	Akcapinar and Bayazit (2018)	Web of Science	Turkish Online Journal of Distance Education
42	Akgun and Demir (2018)	Web of Science	International Journal of Assessment Tools in Education
43	Aybek and Okur (2018)	Web of Science	International Journal of Assessment Tools in Education
44	Afacan Adanir (2019)	Web of Science	Turkish Online Journal of Distance Education
45	Hussain et al. (2019)	Web of Science	International Journal of Emerging Technologies in Learning
46	Guneri and Apaydin (2004)	TR Index Search Tool	Ticaret ve Turizm Egitim Fakultesi Dergisi
47	Tekin (2014)	TR Index Search Tool	Eurasian Journal of Educational Research
48	Akcapinar, Altun, and Askar (2015)*	TR Index Search Tool	Ilkogretim Online
49	Aksoy and Narli (2015)*	TR Index Search Tool	Turkish Journal of Giftedness & Education
50	Dalkilic and Aydin (2017)	TR Index Search Tool	Yuksekogretim ve Bilim Dergisi
51	Ozbay and Ersoy (2017)*	TR Index Search Tool	Gazi Universitesi Gazi Egitim Fakultesi Dergisi
52	Cifci, Kaleli, and Gunal (2018)	TR Index Search Tool	Anadolu Universitesi Egitim Bilimleri Enstitusu Dergisi

*Note. * Produced from the author's thesis.*

Articles indicated by asterisk (*) in Table 3 were not included in the results because they were produced from the thesis work of the authors. As a result, analyzes were conducted over 48 studies.
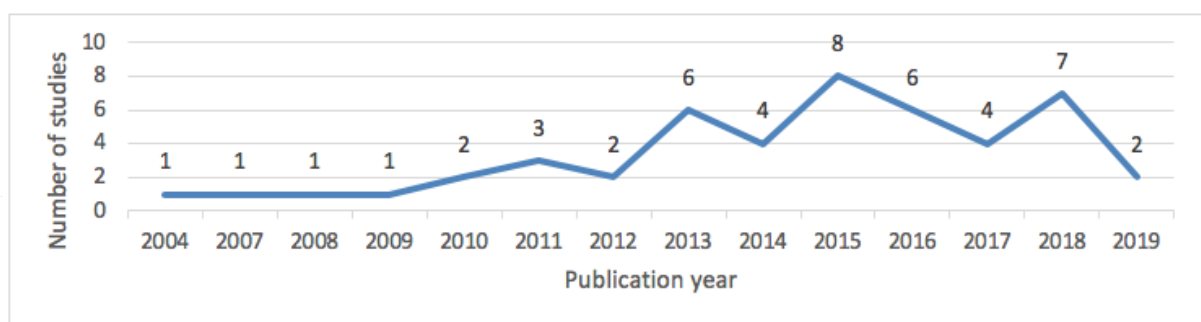

**Figure 1.** Distribution of EDM studies per year of publication

The classification results of the EDM studies according to the data mining tasks are given in Table 4. When Table 4 is analyzed, it is observed that 46.77% of the studies are for prediction task, followed by the tasks of classification, clustering and association rules respectively.

Table 4. Distribution of studies according to data mining tasks

Data mining task	References	%
Classification	Somyurek (2008), Yelegin (2012), Saygili (2013), Sengur (2013), Akcapinar (2014), Aksoy (2014), Ozbay (2015), Uysal (2015), Cebi (2016), Sahin (2018), Guruler et al. (2010), Sohsah et al. (2015), Idil et al. (2016), Ayvaz et al. (2017), Afacan Adanir (2019)	24.19
Prediction	Aydin (2007), Y. Aydogdu (2011), Sengur (2013), Bahadir (2013), Yildiz (2014), Akcapinar (2014), Coskun (2013), Uysal (2015), Yildiz Aybek (2016), Cebi (2016), Barngrover (2017), Yagci (2018), Yorganci (2018), Kentli and Sahin (2011), Turhan et al. (2013), Akcapinar (2015), Demir (2015), Kayri (2015), Sohsah et al. (2015), Bahadir (2016), Kose and Arslan (2016), Ayvaz et al. (2017), Kose and Arslan (2017), Akgun and Demir (2018), Aybek and Okur (2018), Hussain et al. (2019), Guneri and Apaydin (2004), Tekin (2014), Cifci et al. (2018)	46.77
Clustering	Aydin (2007), Ucgun (2009), Birtil (2011), Saygili (2013), Akcapinar (2014), Ozbay (2015), Dogan and Camurcu (2010), Avsar and Yalcin (2015), Baran and Kilic (2015), Onan et al. (2016), Akcapinar and Bayazit (2018), Dalkilic and Aydin (2017)	19.35
Association rules	Ucgun (2009), Y. Aydogdu (2011), Yucel (2012), Hark (2013), Onan et al. (2016), Dalkilic and Aydin (2017)	9.68

*Note. * Some publications are repeated in the table since they involve more than one task.*

The bar graph of the techniques used in EDM studies and the number of studies using these techniques is given in Figure 2. As shown in Figure 2, artificial neural networks is the most commonly used technique in EDM studies (n = 21). It is seen that artificial neural network techniques are used in the researches in predicting student achievement (Akcapinar, 2014; Kose & Arslan, 2016) and classification of words according to usage level (Sohsah et al., 2015). Another technique commonly used in EDM studies is the decision trees. It is seen that decision trees are used in the studies for estimating student achievement (Aydin, 2007), identifying the variables affecting the attitude towards a specific purpose (Idil et al., 2016) and classifying students according to certain characteristics (Guruler et al., 2010; Yelegin, 2012). Other techniques commonly used in EDM research include clustering techniques (Avsar & Yalcin, 2015; Baran & Kilic, 2015; Dogan & Camurcu, 2010), regression (Bahadir, 2016; Coskun, 2013; Turhan et al., 2013), association rules techniques (Hark, 2013; Onan et al., 2016; Yucel, 2012), bayes classifiers (Cebi, 2016; Cifci et al., 2018; Sahin, 2018; Uysal, 2015), and support vector machines (Kentli & Sahin, 2011; Sohsah et al., 2015; Tekin, 2014). In addition to these studies, fuzzy logic (Uysal, 2015), path analysis (Cebi, 2016) and genetic algorithm (Yildiz, 2014) techniques were used.
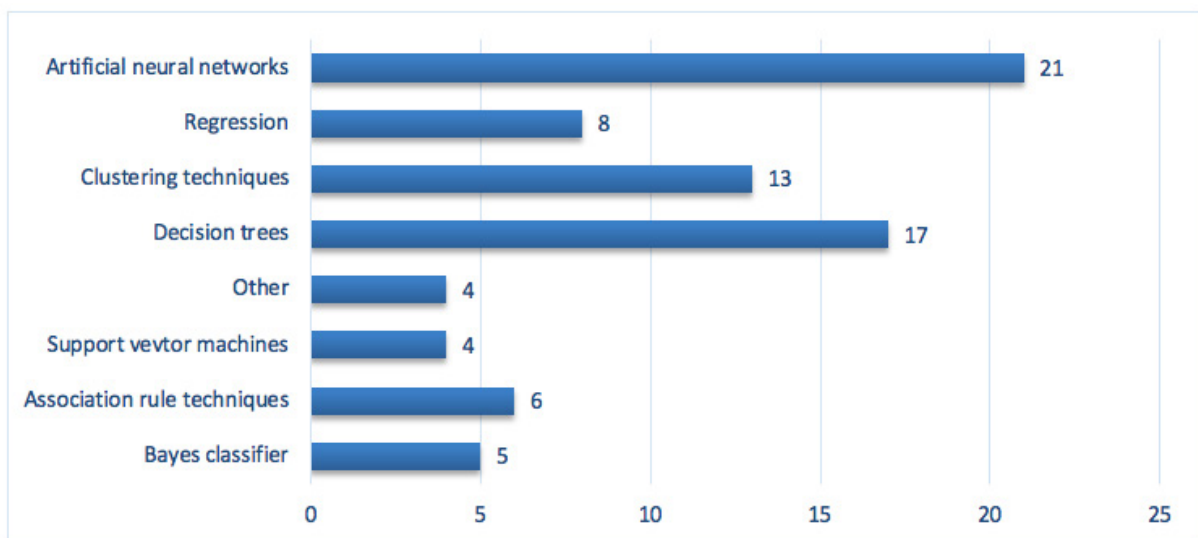
Figure 2. Distribution of the number of studies according to the techniques used in EDM

Classification results according to the topics of the studies related to EDM are given in Table 5. According to the results, publications were mostly aimed to predict student achievement in EDM studies (n = 20).

Achievement prediction studies have tried to predict future success by using the data of the students during the semester, other course achievements or demographic characteristics (Akcapinar, 2014; Akgun & Demir, 2018; Bahadir, 2013; Kayri, 2015; Sengur, 2013; Turhan et al., 2013; Uysal, 2015; Yildiz, 2014). Investigation of e-learning environments was the second most studied topic in EDM studies (n = 7). This topic consist of these studies: determination of the factors determining the efficiency of e-learning environment (Y. Aydogdu, 2011), development of early detection and intervention engine (Sahin, 2018), estimation of students' disorientation levels (Cebi, 2016), estimation of course content for students (Kose & Arslan, 2016, 2017) and the classification of students according to student navigation data (Ozbay, 2015).

Clustering or classification of students according to their success was the most studied another topic (n = 6). This topic includes these research aims: determining successful student characteristics with clustering or classification (Saygili, 2013), clustering by students' achievement (Dogan & Camurcu, 2010; Guruler et al., 2010; Ozbay, 2015) or survey/scale results (Yelegin, 2012) and creating association rules according to successful student characteristics (Onan et al., 2016).

Table 5. Distribution of studies according to topics

Topic of study	Number of studies	%
Achievement/performance prediction	20	39.22
Investigation of e-learning environments	7	13.73
Classification/clustering by success	6	11.76
Classification / clustering according to student characteristics	4	7.84
Investigation of causes of failure	3	5.88
Text mining	3	5.88
Determination of variables that affect attitude	2	3.92
Determination of reasons for absenteeism	1	1.96
Estimation of instructor performance	1	1.96
Determination of effective familial variables on reading skills	1	1.96
Predicting the department that students will prefer	1	1.96
Modeling video navigation	1	1.96
Predicting emotional states with facial recognition	1	1.96
Total	51*	100

*Note. * Since some studies are aimed at more than one purpose, the total value is calculated as 51.*

Classification or clustering according to student characteristics is another topic studied in EDM. This topic includes: clustering according to student profile (Aydin, 2007), student modeling (Somyurek, 2008), determination of the variables that are effective in student modeling (Cebi, 2016) and classification of students according to learning styles (Aksoy, 2014).

According to the results in Table 5, the topics studied in EDM are investigation the causes of students' failures (Baran & Kilic, 2015; Birtil, 2011; Ucgun, 2009), text mining (Afacan Adanir, 2019; Akcapinar, 2015; Sohsah et al., 2015), determination of variables that affect attitude (Hark, 2013; Idil et al., 2016), determination of reasons for absenteeism (Dalkilic & Aydin, 2017), estimation of instructor performance (Cifci et al., 2018), determination of familial variables affecting reading skill (Avsar & Yalcin, 2015), prediction of the department that students will prefer (Coskun, 2013), modeling of video navigations (Akcapinar & Bayazit, 2018) and prediction emotional states with facial recognition and (Ayvaz et al., 2017).

The distribution of the number of studies according to the characteristics of the sample group in EDM is given in Table 6. As seen in Table 6, the sample group in the studies was mostly selected from university students (n = 37). It is thought that researchers prefer to work with university students because the majority of the studies are aimed at prediction of achievement and databases records are used as data source in these studies. The sample group was followed by high school students (n = 5), secondary school students (n = 3) and postgraduate students (n = 1).

Table 6. Distribution of studies according to sample group characteristics

Sample group*	Number of studies	%
University students	37	80.43
High school students	5	10.87
Secondary students	3	6.52
Postgraduate students	1	2.17

*Note. * In some studies were not included in the analysis because they were only data sets (words and face recognition data).*

The distribution of the number of studies according to the data sources used in the EDM is presented in Table 7. When Table 7 is examined, it is seen that the achievement scores are mostly used in the researches as the data source (n = 20). Surveys (n=12), database records (n=10), demographic characteristics (n=7), navigation data (n=5), scales (n = 4) and web-based learning environment (n = 3) are other data sources used.

Table 7. Distribution of studies according to data sources

Data source	Number of studies	%
Achievement scores	20	26.32
Surveys	12	15.79
Database records (Assessment Selection and Placement Center (OSYM) data etc.)	10	13.16
Demographic Characteristics	7	9.21
Navigation data	5	6.58
Scales	4	5.26
Web-based learning environment	3	3.95
Opinion forms	2	2.63
Multiple intelligence scale	2	2.63
Student posts (message-chat tool)	2	2.63
Video interaction data	1	1.32
Disorientation scale	1	1.32
Satisfaction survey	1	1.32
Concealed shapes group test	1	1.32
Process-letter sequence test	1	1.32
Word data set	1	1.32
Prior knowledge level	1	1.32
Learning style inventory	1	1.32
Face recognition data	1	1.32
Total	79*	100

*Note. * Since there may be more than one data source in one study, the total number of studies was calculated as 79.*

The distribution of the number of studies according to the analysis tool used in EDM research is given in Table 8. According to the results, SPSS (n = 8), MATLAB (n = 5), SPSS Clementine (n = 5), WEKA (n = 4) and RapidMiner (n = 3) tools were used in the studies. In addition, data mining tool has been developed in 4 studies. Keras, Neural Connection, R Programlama, SAS Enterprise Manager, SPSS Amos, Tableau, Tensorflow tools were used in only 1 study.

Table 8. Distribution of studies according to the analysis tool used

Analysis tool	Number of studies	%
SPSS	8	22.22
MATLAB	5	13.89
SPSS Clementine	5	13.89
Developed in the study process	4	11.11
WEKA	4	11.11
RapidMiner	3	8.33
Keras	1	2.78
Neural Connection	1	2.78
R Programming	1	2.78
SAS Enterprise Manager	1	2.78
SPSS Amos	1	2.78
Tableau	1	2.78
Tensorflow	1	2.78
Total	36*	100

*Note. * Because in some studies the tool used is not specified or more than one tool is used in a study total value was calculated as 36.*

In the 20 studies (Table 5) for prediction the achievement of students, the results of the analysis on which data are used as input variables for predicting the achievement are presented in Table 9. As the input variable used in the prediction of achievement, it is seen that the midterm exam scores or the previous exam scores of the other courses are the most preferred (n = 13). Demographic characteristics such as gender, age, nationality and region of residence were used in 7 studies. OSYM placement data of university students were used as input variable in 5 researches. The data collected by applying the scale or survey to the students and the system interaction data of the students were used as input variables in 4 studies.

Table 9. Input variables used in predicting achievement

Input variable	Number of studies	%
Midterm or previous exam scores	13	39.39
Demographic characteristics (gender, age, nationality etc.)	7	21.21
OSYM data (high school type, teaching type, high school code, placement score etc.)	5	15.15
Scale or survey data	4	12.12
System interaction values (Duration of use, number of messages, navigation data etc.)	4	12.12
Total	33*	100

*Note. * In some studies, since the input variables were included in more than one class, the total value was calculated as 33.*

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

In this section, the results of the research were discussed in the context of the related literature and suggestions were made for the studies to be carried out regarding EDM. In addition, this section covers the issues that are not included in the results but which have attracted the attention of the researcher.

When the distribution of EDM related studies by years is examined, it is observed that there is limited study for this field. Pena-Ayala (2014) states in his study "EDM is living its spring time and preparing for a hot summer season." However, considering the number of studies conducted for EDM in Turkey, it can be said that this field is a new field of study.

It is seen that the studies carried out in EDM are generally conducted for prediction. In addition, there are studies in which classification, clustering and association rules are performed. Prediction, classification and clustering studies generally focus on model building on a data set and testing the accuracy of the generated model. On the other hand, there was no study in which the model was created dynamically and automatically updated. In the association rules studies, rules were established according to the frequency of coexistence of the data in the data sets. When examining the distribution of analysis tools used in EDM studies, it is observed that Keras (Chollet, 2018), R Programming and Tensorflow (Abadi et al., 2016) tools, which are widely used in Statistics, data mining and machine learning, are used limited research in Turkey and instead these programs ready package programs are preferred. Therefore, it is thought that it would be useful in terms of learning that integration of learning management systems with applications which open source applications developed as an alternative to ready-made package programs used in data mining or machine learning.

In most studies, one or more of the data mining techniques have been used, but the details of the parameters used in these techniques have not been specified. One of the factors determining the performance of a technique is which parameters this technique is used with. Therefore, it is considered that there is a need for studies about parameter optimization in data mining techniques used in EDM studies.

There should be a logical fiction between input and output variables determined for modeling or validation in data mining. For example, Sahin (2018) stated in a study which variables were used in the modeling and why. In some of the studies, it is seen that there is no logical validity to establish the relationship between the content of the questionnaire and student achievement. In particular, it is a big problem that how some input variables used in thesis studies (eg GSM No, TC Identity Number) will contribute to the classification or estimation of student achievement or profile, and how to explain this situation, even if the level of contribution it makes.

When the studies are examined, it is seen that quantitative data (frequency, response to questionnaire, etc.) are used as input variable in clustering, classification and prediction. It is thought that analyzing the content of the shares as well as the sharing frequency of students in educational settings will bring more in-depth results. Therefore, it is recommended that qualitative data should be used in data mining studies and that text mining methods should be employed in the process of preprocessing data. For example, the web-based application developed by Aydogdu and Guyer (2018) for the automatic generation of concept maps can be used as a tool in the preprocessing of textual data since it performs sorting according to the weight value of terms in a document set.

When the techniques used in EDM are examined in Turkey, it is observed that artificial neural networks are used more than other techniques. Similar results are noted in the study by Manjarres et al. (2018) In the literature, artificial neural networks are called "black box" (Benitez, Castro, & Requena, 1997; Olden & Jackson, 2002) and there are different methods for interpreting these networks (de Ona & Garrido, 2014; Olden, Joy, & Death, 2004). When studies on predicting student achievement with artificial neural networks are examined, it is stated in some studies how input variables contribute to student achievement. On the other hand, it is seen that the contribution level of input variables is not explained in the studies or the correlation coefficient between input variable and output variable is interpreted as contribution level. In this case, it is considered that the weight coefficient values in hidden layers in artificial neural networks are not taken into consideration. Finally, in some studies conducted with artificial neural networks, it is seen that midterm exam scores are used as input variables in predicting success. When the results of these studies are examined, it is stated that midterm scores are of great importance as an indicator of the success of the

students. In the case of artificial neural networks, it is normal for the variable to have a significant role if the variable that directly calculates a certain percentage of the output variable is used as the input variable. Therefore it is considered that the use of latent variables or variables that do not have a direct percentage effect on the output variable in these networks is more suitable for the purpose of use of these networks.

In EDM studies, predicting student achievement has been the most studied subject and university students the most selected sample. These results support the results of the review study conducted by Tekin and Oztekin (2018). It is thought that student achievement level and university students are preferred more in the studies because of their ease of access. In addition to these studies, it is recommended to conduct studies at different sample levels and to investigate different variables related to student achievement as well as learning.

LIMITATIONS

In this study, CoHE Thesis Center, Web of Science database and TR Index Search Tool were used according to certain keywords. The studies that exist in the literature but which are not listed in the search results of incomplete or incorrect identification of users, authors or systems in databases are not included in this study.

BIODATA and CONTACT ADDRESSES of AUTHOR
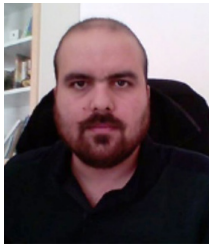
Dr. Seyhmus AYDOGDU is a researcher at Nevsehir Haci Bektas Veli University, Department of Computer Education and Instructional Technologies. He gained his MSc and Ph.D. degrees in Computer Education and Instructional Technologies from Gazi University at 2012 and 2016, respectively. He worked as expert programmer at Ahmet Yesevi University TURTEP between 2010 and 2016. His academic researches focus on educational data mining, educational software development, machine learning techniques in educational environments and teaching programming.

Seyhmus AYDOGDU
Department of Computer Education and Instructional Technologies, Faculty of Education
Address: Nevsehir Haci Bektas Veli University, 50300, Nevsehir, TURKEY
Phone: +90 3842281000 - 21051,
E-mail: saydogdu@nevsehir.edu.tr, aydogduseyhmus@gmail.com

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Abu Tair, M. M., & El-Halees, A. M. (2012). Mining educational data to improve students' performance: a case study. *Mining educational data to improve students' performance: a case study, 2*(2).

Afacan Adanir, G. (2019). DETECTING TOPICS OF CHAT DISCUSSIONS IN A COMPUTER SUPPORTED COLLABORATIVE LEARNING (CSCL) ENVIRONMENT. *Turkish Online Journal of Distance Education, 20*(1), 96-114.

Akcapinar, G. (2015). How automated feedback through text mining changes plagiaristic behavior in online assignments. *Computers & Education, 87*, 123-130. doi:10.1016/j.compedu.2015.04.007

Akcapinar, G., & Bayazit, A. (2018). INVESTIGATING VIDEO VIEWING BEHAVIORS OF STUDENTS WITH DIFFERENT LEARNING APPROACHES USING VIDEO ANALYTICS. *Turkish Online Journal of Distance Education, 19*(4), 116-125.

Akcapinar, G. (2014). Cevrimici ogrenme ortamindaki etkilesim verilerine gore ogrencilerin akademik performanslarinin veri madenciligi yaklasimi ile modellenmesi. (Doctoral dissertation), Hacettepe Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Akcapinar, G., Altun, A., & Askar, P. (2015). Modeling students' academic performance based on their interactions in an online learning environment. Ilkogretim Online, 14(3).

Akgun, E., & Demir, M. (2018). Modeling Course Achievements of Elementary Education Teacher Candidates with Artificial Neural Networks. *International Journal of Assessment Tools in Education, 5*(3), 491-509. doi:10.21449/ijate.444073

Aksoy, E. (2014). *Matematik alaninda ustun yetenekli ve zekali ogrencilerin bazi degiskenler acisindan veri madenciligi ile belirlenmesi.* (Master's thesis), Dokuz Eylul Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Aksoy, E., & Narli, S. (2015). An Examination of Mathematically Gifted Students' Learning Styles by Decision Trees. *Turkish Journal of Giftedness & Education, 5*(2).

Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., . . . Olatunji, S. O. (2017, 30 April-3 May 2017). *Student performance prediction using Support Vector Machine and K-Nearest Neighbor.* Paper presented at the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics, 37*, 13-49. doi:https://doi.org/10.1016/j.tele.2019.01.007

Avlijas, G., Heleta, M., & Avlijas, R. (2016). *A guide for association rule mining in moodle course management system.* Paper presented at the Proceedings of the International Scientific Conference-Sinteza 2016,(Belgrade, Serbia: Singidunum University).

Avsar, A. S., & Yalcin, S. (2015). Determining the Parental Variables That Explain Students' Reading Success by Using CHAID Analysis. *EGITIM VE BILIM-EDUCATION AND SCIENCE, 40*(179), 1-9.

Aybek, H. S. Y., & Okur, M. R. (2018). Predicting Achievement with Artificial Neural Networks: The Case of Anadolu University Open Education System. *International Journal of Assessment Tools in Education, 5*(3), 474-490. doi:10.21449/ijate.435507

Aydin, S. (2007). *Veri madenciligi ve Anadolu Universitesi uzaktan egitim sisteminde bir uygulama.* (Doctoral dissertation), Anadolu Universitesi, Sosyal Bilimler Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Aydogdu, Y., & Tanrikulu, Z. (2013). Corporate E-Learning Success Model Development by Using Data Mining Methodologies. *EGITIM VE BILIM-EDUCATION AND SCIENCE, 38*(170), 95-111.

Aydogdu, S., & Guyer, T. (2018). Development of a Web-Based Application for the Automatic Creation of Concept Maps. *Bartin Universitesi Egitim Fakultesi Dergisi, 7*(2), 349-364.

Aydogdu, Y. (2011). *Evaluating e-learning environment by using data mining techniques.* (Master's thesis), Bogazici Universitesi, Sosyal Bilimler Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Ayvaz, U., Guruler, H., & Devrim, M. O. (2017). USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS. *Information Technologies and Learning Tools, 60*(4), 95-104. doi:10.33407/itlt.v60i4.1743

Bahadir, E. (2016). Using Neural Network and Logistic Regression Analysis to Predict Prospective Mathematics Teachers' Academic Success upon Entering Graduate Education. *Educational Sciences-Theory & Practice, 16*(3), 943-964. doi:10.12738/estp.2016.3.0214

Bahadir, E. (2013). *Yapay sinir aglari ve lojistik regresyon analizi yaklasimlari ile ogretmen adaylarinin akademik basarilarinin tahmini.* (Doctoral dissertation), Marmara Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Baker, R. (2010). Data mining for education. *International encyclopedia of education, 7*(3), 112-118.

Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.3417.*

Baran, B., & Kilic, E. (2015). Applying The CHAID Algorithm to Analyze How Achievement is Influenced by University Students' Demographics, Study Habits, and Technology Familiarity. *Educational Technology & Society, 18*(2), 323-335.

Barngrover, M. K. (2017). *Description and prediction: Knowledge discovery in university databases.* (Master's thesis), Bogazici Universitesi, Sosyal Bilimler Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Benitez, J. M., Castro, J. L., & Requena, I. J. I. T. o. n. n. (1997). Are artificial neural networks black boxes? *, 8*(5), 1156-1164.

Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*: John Wiley & Sons.

Bhardwaj, B. K., & Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. *arXiv preprint arXiv:1201.3418.*

Birtil, F. S. (2011). *Kiz meslek lisesi ogrencilerinin akademik basarisizlik nedenlerinin veri madenciligi teknigi ile analizi.* (Master's thesis), Afyon Kocatepe Universitesi, Fen Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Bousbia, N., & Belamri, I. (2014). Which Contribution Does EDM Provide to Computer-Based Learning Environments? In A. Pena-Ayala (Ed.), *Educational Data Mining: Applications and Trends* (pp. 3-28). Cham: Springer International Publishing.

Bucos, M., & Dragulescu, B. (2018). Predicting Student Success Using Data Generated in Traditional Educational Environments. *Tem Journal-Technology Education Management Informatics, 7*(3), 617-625. doi:10.18421/tem73-19

Chollet, F. (2018). Keras. Retrieved from https://keras.io

Coskun, M. (2013). *A web based multi-criteria decision support system for department selection process of vocational high school students.* (Master's thesis), Bogazici Universitesi, Sosyal Bilimler Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Cebi, A. (2016). *Olasiliga dayali uyarlanabilir hiper ortamlarda etkili olan degiskenlerin modellenmesi.* (Doctoral dissertation), Gazi Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Cifci, F., Kaleli, C., & Gunal, S. (2018). Oznitelik Secme ve Makine Ogrenmesi Yontemleriyle Egitmen Performansinin Tahmin Edilmesi. *Anadolu Universitesi Egitim Bilimleri Enstitusu Dergisi, 8*(2), 419-440.

Dalkilic, F., & Aydin, O. (2017). Dokuz Eylul Universitesi Iktisadi ve Idari Bilimler Fakultesi Ogrencilerinin Devamsizlik Davranislarini Etkileyen Faktorler. *Journal of Higher Education & Science/Yuksekogretim ve Bilim Dergisi, 7*(3).

de Ona, J., & Garrido, C. (2014). Extracting the contribution of independent variables in neural network models: a new approach to handle instability. *Neural Computing and Applications, 25*(3), 859-869. doi:10.1007/s00521-014-1573-5

Demir, M. (2015). Predicting Pre-service Classroom Teachers' Civil Servant Recruitment Examination's Educational Sciences Test Scores Using Artificial Neural Networks. *Educational Sciences-Theory & Practice, 15*(5), 1169-1177.

Dogan, B., & Camurcu, A. Y. (2010). Visual Clustering of Multidimensional Educational Data From an Intelligent Tutoring System. *Computer Applications in Engineering Education, 18*(2), 375-382. doi:10.1002/cae.20272

Dutt, A., Aghabozrgi, S., Ismail, M. A. B., & Mahroeian, H. (2015). Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics Engineering, 5*(2), 112.

Francis, B. K., & Babu, S. S. (2019). Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. *Journal of Medical Systems, 43*(6), 15. doi:10.1007/s10916-019-1295-4

Guruler, H., Istanbullu, A., & Karahasan, M. (2010). A new student performance analysing system using knowledge discovery in higher educational databases. *Computers & Education, 55*(1), 247-254. doi:10.1016/j.compedu.2010.01.010

Guneri, N., & Apaydin, A. (2004). Ogrenci basarilarinin siniflandirilmasinda lojistik regresyon analizi ve sinir aglari yaklasimi. *Ticaret ve Turizm Egitim Fakultesi Dergisi, 1*, 170-188.

Hark, C. (2013). Ogrencilerin akilli tahtaya iliskin tutumlarinin incelenmesine yonelik bir veri madenciligi uygulamasi. (Master's thesis), Firat Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Hussain, S., Muhsin, Z. F., Salal, Y. K., Theodorou, P., Kurtoglu, F., & Hazarika, G. C. (2019). Prediction Model on Student Performance based on Internal Assessment using Deep Learning. *International Journal of Emerging Technologies in Learning, 14*(8), 4-22. doi:10.3991/ijet.v14i08.10001

Idil, F. H., Narli, S., & Aksoy, E. (2016). Using Data Mining Techniques Examination of the Middle School Students' Attitude towards Mathematics in the Context of Some Variables. *International Journal of Education in Mathematics Science and Technology, 4*(3), 210-228. doi:10.18404/ijemst.02496

Kamisli Ozturk, Z., Erzurum Cicek, Z. I., & Ergul, Z. (2017). Sentiment analysis: An application to Anadolu University. *Acta Physica Polonica A*, 132(3), 753–755. https://doi.org/10.12693/APhysPolA.132.753.

Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. *Procedia Computer Science, 57*, 500-508. doi:https://doi.org/10.1016/j.procs.2015.07.372

Kayri, M. (2015). An Intelligent Approach to Educational Data: Performance Comparison of the Multilayer Perceptron and the Radial Basis Function Artificial Neural Networks. *Educational Sciences-Theory & Practice, 15*(5), 1247-1255.

Kentli, F. D., & Sahin, Y. (2011). An SVM approach to predict student performance in manufacturing processes course. *Energy Education Science and Technology Part B-Social and Educational Studies, 3*(4), 535-544.

Kose, U., & Arslan, A. (2016). Intelligent E-Learning System for Improving Students' Academic Achievements in Computer Programming Courses. *International Journal of Engineering Education, 32*(1), 185-198.

Kose, U., & Arslan, A. (2017). Optimization of Self-Learning in Computer Engineering Courses: An Intelligent Software System Supported by Artificial Neural Network and Vortex Optimization Algorithm. *Computer Applications in Engineering Education, 25*(1), 142-156. doi:10.1002/cae.21787

Lin, C. F., Yeh, Y.-c., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers & Education, 68*, 199-210. doi:https://doi.org/10.1016/j.compedu.2013.05.009

Man, M., Abu Bakar, W. A. W., & Sabri, I. A. A. (2018). An Association Rule on eDisiplin Case Study: An Educational Data Mining Approach. *Advanced Science Letters, 24*(3), 1872-1875. doi:10.1166/as1.2018.11179

Manjarres, A. V., Sandoval, L. G. M., & Suarez, M. J. S. (2018). Data mining techniques applied in educational environments: Literature review. *Digital Education Review*(33), 235-266.

Mohamad, S. K., & Tasir, Z. (2013). Educational Data Mining: A Review. *Procedia - Social and Behavioral Sciences, 97*, 320-324. doi:https://doi.org/10.1016/j.sbspro.2013.10.240

Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling, 154*(1-2), 135-150.

Olden, J. D., Joy, M. K., & Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling, 178*(3), 389-397. doi:https://doi.org/10.1016/j.ecolmodel.2004.03.013

Onan, A., Bal, V., & Bayam, B. Y. (2016). The Use of Data Mining for Strategic Management: A Case Study on Mining Association Rules in Student Information System. *Croatian Journal of Education-Hrvatski Casopis Za Odgoj I Obrazovanje, 18*(1), 41-70.

Ozbay, O. (2015). Ogretim yonetim sistemi uzerinde universite (lisans) duzeyindeki ogrenci hareketliliginin veri madenciligi yontemleriyle analizi. (Master's thesis), Baskent Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Ozbay, O., & Ersoy, H. (2017). Ogrenme Yonetim Sistemi Uzerindeki Ogrenci Hareketliliginin Veri Madenciligi Yontemleriyle Analizi. *Gazi Universitesi Gazi Egitim Fakultesi Dergisi, 37*(2), 523-558.

Pena-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications, 41*(4, Part 1), 1432-1462. doi:https://doi.org/10.1016/j.eswa.2013.08.042

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*(1), 135-146. doi:https://doi.org/10.1016/j.eswa.2006.04.005

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), 40*(6), 601-618. doi:10.1109/TSMCC.2010.2053532

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12-27.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (2010). *Handbook of educational data mining*: CRC press.

Saygili, A. (2013). *Veri madenciligi ile muhendislik fakultesi ogrencilerinin okul basarilarinin analizi.* (Master's thesis), Yildiz Teknik Universitesi, Fen Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Sohsah, G. N., Unal, M. E., & Guzey, O. (2015). Classification of word levels with usage frequency, expert opinions and machine learning. *British Journal of Educational Technology, 46*(5), 1097-1101. doi:10.1111/bjet.12338

Somyurek, S. (2008). *Uyarlanabilir egitsel web ortamlarinin ogrencilerin akademik basarisina ve gezinmesine etkisi.* (Doctoral dissertation), Gazi Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Sahin, M. (2018). *E-ogrenme ortamlarina yonelik ogrenme analitiklerine dayali mudahale motoru tasarimi ve gelistirilmesi.* (Doctoral dissertation), Hacettepe Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Sengur, D. (2013). Ogrencilerin akademik basarilarinin veri madenciligi metotlari ile tahmini. (Master's thesis), Firat Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Tekin, A. (2014). Early prediction of students' grade point averages at graduation: A data mining approach. *Eurasian Journal of Educational Research, 54*, 207-226.

Tekin, A., & Oztekin, Z. (2018). Egitsel veri madenciligi ile ilgili 2006-2016 yillari arasinda yapilan calismalarin incelenmesi. *Egitim Teknolojisi Kuram ve Uygulama, 8*(2), 108-124.

Turhan, K., Kurt, B., & Engin, Y. Z. (2013). Estimation of Student Success with Artificial Neural Networks. *EGITIM VE BILIM-EDUCATION AND SCIENCE, 38*(170), 112-120.

Uysal, H. (2015). *BOTE bolumu ogrencilerinin akademik basarilarini etkileyen faktorlerin belirlenmesi ve akademik basari tahmin modelinin gelistirilmesi.* (Master's thesis), Canakkale Onsekiz Mart Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Ucgun, K. (2009). *Orta ogretim okullari icin ogrenci otomasyonu tasarimi ve ogrenci verileri uzerine veri madenciligi uygulamalari.* (Master's thesis), Marmara Universitesi, Fen Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Yagci, A. (2018). *Mesleki ve teknik lise ogrencilerinin fen dersleri (fizik- kimya- biyoloji) basarilarinin yapay sinir aglari ile tahmini ve basarisizlik icin alinacak tedbirler (Turkiye-Malezya karsilastirmasi).* (Master's thesis), Karamanoglu Mehmetbey Universitesi, Fen Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Yang, F., & Li, F. W. B. (2018). Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Computers & Education, 123*, 97-108. doi:https://doi.org/10.1016/j.compedu.2018.04.006

Yelegin, A. (2012). *Mesleki egitimde ogrenci altyapisinin ogrenci egitim basarisina etkisinin veri madenciligi yontemleriyle ortaya cikarilmasi.* (Master's thesis), Beykent Universitesi, Fen Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Yildiz Aybek, H. S. (2016). Ogrenci basarisinin yapay sinir aglari ile kestirilmesi: Anadolu Universitesi Acikogretim Sistemi ornegi. (Master's thesis), Anadolu Universitesi, Sosyal Bilimler Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Yildiz, O. (2014). *Makine ogrenmesi ile uzaktan egitim ogrencilerinin performanslarinin degerlendirilmesi.* (Doctoral dissertation), Istanbul Universitesi, Fen Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Yorganci, N. (2018). Ogretmenlik meslegine karsi tutum ile akademik basari arasindaki iliskinin yapay sinir aglari ile analizi. (Master's thesis), Burdur Mehmet Akif Ersoy Universitesi, Egitim Bilimleri Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education, 29*, 23-30. doi:https://doi.org/10.1016/j.iheduc.2015.11.003

Yucel, O. (2012). *Development of a data mining software on higher educational data.* (Master's thesis), Bogazici Universitesi, Sosyal Bilimler Enstitusu. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/