

Data Mining Studies in Education: Literature Review For The Years 2014-2020

Zehra Bilici*, Durmuş Özdemir**

Makale Geliş Tarihi: 30/12/2020

Makale Kabul Tarihi: 03/04/2021

DOI: 10.35675/befdergi.849973

Abstract


Data mining is one of the important and beneficial technological developments in education and its usage area is becoming widespread day by day as it includes applications that contribute positively to teaching activities. By making raw data in the field of education meaningful using data mining techniques, teaching activities can be made more effective and efficient. Studies carried out in the field of education between 2014-2020 with data mining methods were scanned from the "Science Direct" database. As a result of scanning studies, 60 papers were found to be directly related to data mining in education. The studies include issues such as the development of e-learning systems, pedagogical support, clustering of educational data, and student performance predictions. These selected articles were analyzed in terms of purpose, application area, method, and contribution to the literature. This study aims to group the studies conducted in the field of education using the data mining method under certain headings, evaluate the methods and goals and present the need in this field to the researchers who will work in this field.


Keywords: *Data mining in education, data mining technologies, technologies for teaching, trends in education*

Eğitimde Veri Madenciliği Çalışmaları: 2014-2020 Yıllarına Ait Literatür Taraması

Öz

Veri madenciliği eğitimde önemli ve faydalı teknolojik gelişmelerden biridir ve öğretim faaliyetlerine olumlu yönde katkı sağlayan uygulamaları içerdiği için kullanım alanı gün geçtikçe yaygınlaşmaktadır. Veri madenciliği teknikleri kullanılarak eğitim alanındaki ham verilerin anlamlı hale getirilmesi ile öğretim etkinliklerinin daha etkin ve verimli hale getirilmesi mümkündür. Veri madenciliği yöntemleriyle 2014-2020 yılları arasında eğitim alanında yapılan çalışmalar "Science Direct" veri tabanından taranmıştır. Tarama çalışmaları

* Kütahya Dumlupınar University, Department of Computer Engineering, Kütahya, Turkey, zehra.bilici@ogr.dpu.edu.tr, ORCID: 0000-0002-5417-428X 

** Kütahya Dumlupınar University, Department of Computer Engineering, Kütahya, Turkey, durmus.ozdemir@dpu.edu.tr, ORCID: 0000-0002-9543-4076 

Kaynak Gösterme: Bilici, Z. & Özdemir, D. (2022). Data Mining Studies in Education: Literature Review For The Years 2014-2020. *Bayburt Eğitim Fakültesi Dergisi*, 17(33), 342-376.

sonucunda, 60 araştırmanın doğrudan eğitimde veri madenciliği ile ilişkili olduğu tespit edilmiştir. Seçilen araştırmalar, makalenin amacı, uygulama alanı ve örneklem, metot ve yöntemi, literatüre katkısı şeklinde tasnif edilerek sunulmuştur. Çalışmalar e-öğrenme sistemlerinin geliştirmesi, pedagojik destek, eğitim verilerinin kümelenmesi, öğrenci performans tahminleri konu başlıkları altında kategorize edilmiştir. Bu çalışmada, veri madenciliği yöntemi kullanılarak eğitim alanında yapılan çalışmalarını belirli başlıklar altında gruplamak, yöntemlerini ve amaçlarını belirlemek ve bu alanda çalışacak olan kişilere alandaki ihtiyacı göstermek amaçlanmıştır.

Anahtar Kelimeler: Eğitimde veri madenciliği, eğitimde yeni yaklaşımlar, öğretim teknolojileri, veri madenciliği teknolojileri

Introduction

The widespread use of developments in instructional technology and cloud systems in education areas and making the data in this field meaningful make education more effective and meaningful. Data mining is important to make the data meaningful, contribute to the people working in education, and provide accurate information for the students (Ahmed et al., 2016; Koedinger et al., 2015). The fact that the cost of computer systems is declining day by day and that their output is increasing makes it possible to store greater volumes of data on the computer. Such conditions allow the data to exceed very broad dimensions. Techniques that process large quantities of data and make them accessible are gaining importance with the advancement of technologies today (Romero & Ventura, 2013). The conversion of raw data into information or meaningful can be done by data mining (Thuraisingham, 2003). Data mining is the process of discovering patterns and trends stored in large sets of data used in these cases. Data mining is also a technique of data analysis that helps to find the connection between them by analyzing relationships within a very large amount of data. Data mining occurs by taking preliminary information about the field of the study and making use of decision-making systems after discovering the information available in databases (Viloria et al., 2019).

Data mining ensures that confidential information is extracted within cloud-based database systems consisting of large datasets. This process can be done using statistics, mathematics disciplines, modeling techniques, database technology, and various computer programs (Romero & Ventura, 2013). Some of the application fields in which data mining is used are health, business, public, market research, banking, finance and exchange, internet, risk analysis, corporate resource usage, trade, consumer analysis, shopping, banking, telecom, medicine, science, engineering, insurance, and education (Agarwal et al., 2012).

There are major technical advances in the field of education today. Huge piles of data have arisen in the field of education with the advent of technology and these data mining issues have been attempted to be solved (Baker, 2010). In the field of education, there is a lot of information that can be useful for data mining studies,

students, teachers, administrative staff, educational institutions, and distance education (Balaman, 2020). Therefore, the effect of data mining on the development of the future education system is great (Chalaris et al., 2014). The results of data mining will help students direct the learning process, help educators to develop their teaching practices, and assist in the administrative tasks of the management process.

The knowledge created by data mining activities is generally aimed at improving teaching methods, such as the evaluation, monitoring, and personalization of the teaching process. For example, educators can monitor students' learning levels and regularly re-adapt the teaching cycle. Thus, educators can adjust their methods and teaching practices, and students can use assessment tools to test their knowledge.

In this study, a literature review was conducted to determine how data mining methods are used in the field of education and in which application areas. The aim of the study is to contribute to people who want to practice in the field of education by using data mining methods later on. Furthermore, the subjects needed in the literature are aimed to determine by presenting current application areas, methods, and study objectives. For this reason, in the last 6 years, the work carried out in this area was reviewed and a perspective was given to promote the understanding of needs.

What is the Data Mining?

Information systems at first focused on simple file structures in the process of data processing, while with the emergence of advanced database systems, data warehouse was developed and the concept of data mining emerged. Data mining is an interdisciplinary area in which computer science, machine learning, management of databases, mathematical algorithms, and statistics are combined (Liao, 2003). Furthermore, data mining can be referred to as a technique of data analysis that helps to find the correlation between them by analyzing the relationships within a vast volume of data and allows hidden information within database systems to be retrieved. Data mining is not a solution on its own, but an instrument that assists the decision-making process in reaching a solution and provides the knowledge required to solve the problem.

Data mining has several different definitions: Data mining is essentially a method of computer-aided knowledge processing (Vranic et al., 2007). The method of discovering useful information by analyzing data from vast quantities of data is data mining (Baker, 2010). Data mining analyzes datasets to define unsuspecting relationships within the dataset and summarize the data in a new way in a useful and relevant manner to the data owner (Romero & Ventura, 2020). Data mining; helps the analyst to find the patterns and relationships between the data emerging in the business phase. Data mining processes require different cross-field work. These fields as shown in Figure 1, are disciplines such as database systems, data visuality, artificial neural networks, statistics, artificial learning.



Figure 1. Disciplines related to data mining

Data Mining and Application Fields

Limiting the use of the data mining method to specific areas of application is not the right approach. Because data mining can be used in almost all fields where data is used. Health, industry, engineering, marketing, banking, and education are the main application fields where data mining is heavily used (Bhullar & Kaur, 2012). One of the most important application fields where data mining is used is health studies. Studies in this field have an important role in the development of drugs, detection of drug effects, prediction of patient test results, pre-diagnosis, and treatment of diseases (Chen et al., 2017).

In the field of industry and engineering, data mining is used to understand the data obtained from the computer environment, to control production processes, to perform quality control analyses, to remove factors and rules that affect system performance. Y. Li & Zhai (2018) used the visualization method and feature extraction method of data mining to understand the underlying dynamics of building energy consumption. In the public sphere, data mining methods are used in the correct use of institutional resources, ensuring public security, and predicting security problems (Xu et al., 2019).

One of the most common application fields of data mining is marketing. Data mining is used in many applications such as making sales forecasting, managing customer relations, doing customer analysis, increasing profitability rate (Amado et al., 2018). Data mining in the field of banking, finance, and the stock market has a wide range of applications. Data mining is used in the evaluation of credit card and credit claims, risk analysis, risk management, forecasting stock prices, and modeling investments.

Due to the development of computer technologies and increased internet usage, data mining in the internet field was a wide range of applications. Data mining is used in fields such as identifying user profiles via social media, identifying malicious users, and ensuring that web pages are configured according to user details (Romero & Ventura, 2007).

Data Mining in Education

In this part of the study, we presented in detail the use and related methods of data mining in the field of education. Data mining is used in many different fields of education. Data mining is used to analyze student data, identify the causes of student achievement and failure, increase student achievement, detect disruptions in educational environments, and create more effective educational environments (Vranic et al., 2007). Educational Data Mining (EDM) can be described as the application of conventional data mining techniques to the analysis of training data aimed at solving educational problems (Fernandes et al., 2019). The process of data mining in education is illustrated in Figure 2.

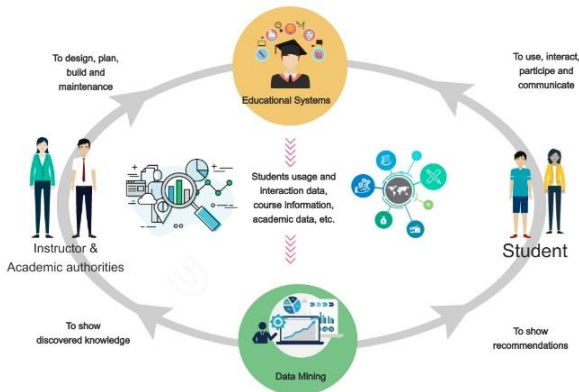


Figure 2. Educational data mining

Some EDM applications include the development of e-learning systems (Burgos et al., 2018), pedagogical support Ahmed et al. (2016), clustering of education data Chakraborty et al. (2016) predictions for student results Kaur et al. (2015), the effects of technology for early warning and telehealth on nurses Kaur et al. (2015), summarize the strategic requirements for an urban design project in architectural fiction Valls et al. (2018), to make education more technological, to show how students share their ideas with twitter and what category of comments are made Patil & Kulkarni, (2018), identifying students who are likely to fail at an early stage Costa et al. (2017), spreads over a wide area such as. As a result of the researches, the studies usually focus on student performance predictions.

Methods of Educational Data Mining

When data mining methods used in educational applications in the literature are examined, there are different types of classifications. One of the forms of classification is based on whether the methods are parametric. The conditions of various assumptions about the properties studied in parametric models are sought. Therefore, the use of Parametric models in real life is very limited. Nonparametric models are more suitable for data mining because these models create a model based on data.

Non-parametric techniques are classified as follows;

- Neural networks (J. Guo et al., 2020)
- Decision trees (Tayefi et al., 2017)
- Genetic algorithms (Hong et al., 2018)

Models used in data mining are examined under two main headings according to another classification rule.

Predictive Models: In these models, a model is developed based on the data whose results are known, and using this model is aimed to estimate the resulting values of the datasets whose results are unknown. If the predicted variable is a numerical variable, problems with estimation are called problems with regression, and they are categorically called problems with classification (Bhullar & Kaur, 2012).

Descriptive Models: These types of models provide identification of models in existing data that can be used to guide decision-making.

According to the functions of data mining models;

- Classification and regression (Yucel et al., 2020)
- Clustering (H. Wang et al., 2020)
- Association rules (Shabtay et al., 2020)
- Sequential patterns (Saleti & Subramanyam, 2019)

Methods and techniques in data mining, such as classification, clustering, and union rules, are also used in educational data mining. The most widely used method among these is the classification method. In the classification method, the dataset is divided into two parts as training and testing. A model is created by analyzing training data classification algorithms. The performance of the model is checked on the test

data at the next point, and the results are measured. The process followed in the classification method is shown in Figure 3.

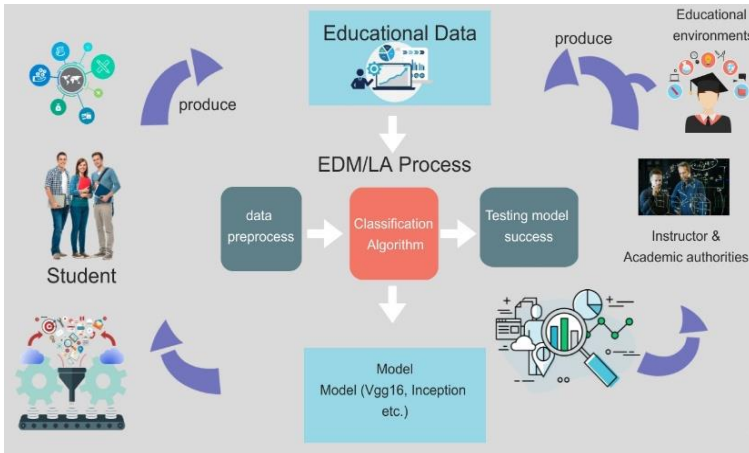


Figure 3. Educational data mining methods

A model is developed by evaluating algorithms for the classification of educational data, and students whose characteristics (social, personal, and academic) are known are compared with test data. To determine the success of the model, the actual class of each sample in the test data and the class predicted by the model are compared. The second of the educational data mining methods is the "clustering method", which, in addition to classification, replaces two sets of data with a single dataset. The current dataset is analyzed with a selected clustering algorithm, and the dataset is divided into several sets based on a model entirely generated by the algorithm. In this method, using similarities and differences in the dataset, algorithms perform clustering operations. Models and libraries used in data mining studies in the field of education are as follows:

Artificial Neural Network Models: Convolutional Neural Network-CNN (Zhao et al., 2020), Recurrent Neural Network-RNN (Okubo et al., 2017), Long Short-Term Memory-LSTM (X. Li & Wu, 2015), Restricted Boltzmann Machine-RBM (B. Guo et al., 2015), Deep Belief Network-DBN (Zhang et al., 2020), Auto Encoder (Converse et al., 2019).

Deep Learning Algorithms: LeNet (Guan et al., 2020), AlexNet (Hernández-Blanco et al., 2019), VggNet (Kim et al., 2016), GoogleNet (L. Wang et al., 2017).

Deep Learning Libraries: TensorFlow (Fok et al., 2018), Caffe (Cabada et al., 2018), Theano (B. Guo et al., 2015), Torch (Pandey & Bhardwaj, 2012), DeepLearning4j (Lang et al., 2019), Keras (Aydoğdu, 2020), Cognitive Network Toolkit-CNTK (Jiang & Li, 2020), Lasagne (Notten, 2014).

Method

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

A systematic review approach was used in this study, which was designed to analyze the studies carried out using the data mining method in the field of education. The systematic review is used to identify, select, and critically evaluate relevant research and analyze the studies included in the review (Thuraisingham, 2003). The study aims to group the studies conducted in the field of education under specific headings, using the method of data mining, to classify their methods and goals, and to direct the people who will work in this field. Another aim of this study is to determine the subjects needed in the literature by presenting current application areas, methods, and working objectives. The keywords "education" and "data mining" were used for this literature study.

The study consists of research, selection, and finalization stages. Progress was made according to the stages schematized in Figure.3. Firstly, a scan was carried out in the 'Science Direct' database based on the words "education" and "data mining". The studies carried out in the field of data mining were examined specifically to answer the question "Which specific area and in how it is implemented?" As a result of stage-2 research, 60 articles were selected from 163 articles that were appropriate for the target. Stage-3 is presented by grouping and table representation by the content analysis method of articles. The aim of the evaluated articles, the type of study, application areas, methods, and the process of converting the results into a summarized table are shown in Figure-4.

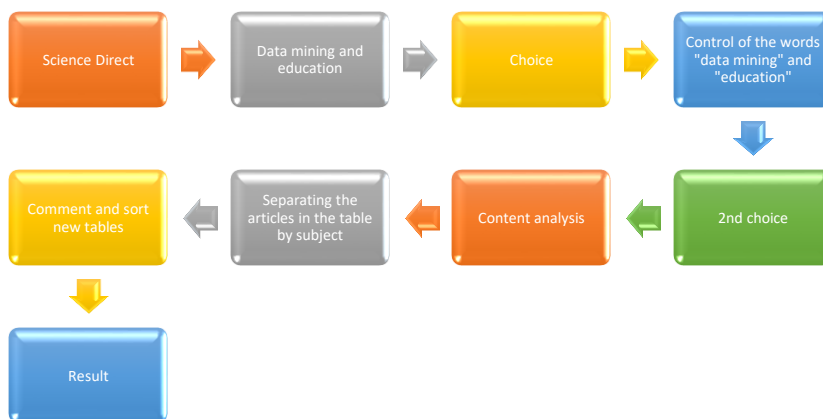


Figure 4. Research process

Literatur Survey

Studies aimed at predicting student achievement status are presented below when the articles made in the field of education using the data mining approach between 2014 and 2020 are analyzed.

Shahiri & Husain (2015) discussed the subject of an evaluation of student performance with data mining techniques to search the literature. Ahmed et al. (2016) examined data mining predictions for the prediction of student performance in order to increase the quality of education, 4 different data mining classification methods were studied for this study and the best method was chosen. In a study by Asif et al. (2017), studied students for 4 years to predict the future achievements of undergraduates and compared the estimated results. The aim of Campagni et al. (2015) is to predict the performance of students after finishing university. Miguéis et al. (2018) estimated the performance of the students with data from approximately 2500 students studying engineering in Europe between 2003 and 2015 and as a result of this study, success is 95%. Burgos et al. (2018) successfully estimated students' academic achievement using data from former students in the 2015 and 2016 semesters in public schools in the Federal state of Brazil. Costa et al. (2017) studied distance education and campus students at The Public University of Brazil, and the study aimed to predict students' performance and develop educational techniques for failed students. In their research, Oeda and Hashimoto (2017) aimed to infer students who could not keep up with programming lessons, rather than predict students' performance from their daily data. In particular, they proposed a method for predicting dropouts using outlier value detection to cluster data with unsupervised learning. Reimer et al. (2015) model the performance of primary and secondary school students in mathematics as a stochastic process to predict eighth-grade performance. The study of Badr et al. (2016), discusses the building of a model to predict the performance of students in a programming course based on their grades in courses in other subjects. Lara et al. (2014) predict student performance to provide teachers with inferences from past student performance. Sandoval et al. (2018) present a prediction model based on predicting early which students attending large classes (with more than 50 enrollments) are at risk of failing a course.

In Table 1, studies aimed at predicting student success are presented as the type, purpose, application field, method, and contribution to the literature. 20 of the 60 types of research were based on studies on predicting student performance. When the studies are examined, 12 of the studies on the prediction of student performance are research articles, 3 are conference papers, and 5 are book type researches.

Table 1.

Prediction of Student Performance

The People Who Made the Work	Type of Study	Purpose of the study / Aim (Objective)	Application Area and Sample	Metod	Conclusion- Positive/Negative-Contribution to The Field Literature
Sen & Ucar, (2012)	Article	To compare student achievements.	Undergraduate students	Comparison	Positive- 97% gave correct results.
Lara et al., (2014)	Article	Analyzing data from students' interaction with e-learning environments.	Students and lessons	Analysis	Positive-Best results have been obtained.
Reamer et al., (2015)	Article	Predicting the performance of elementary and secondary school students in mathematics for eighth grade.	Elementary and secondary school students	Evaluation and grouping	Positive-Student performance has been estimated.
Gupta et al., (2015)	Conference paper	To show the importance of information evaluation.	Higher education institutions	Evaluation	Positive- The importance of information evaluation has been demonstrated.
Campagni et al., (2015)	Article	It offers a data mining methodology for analyzing the careers of university graduate students.	College students	Analysis	Positive- The predictions made in students' careers are correct.
Shahiri & Husain, (2015)	Conference paper	Predicting students' performance.	Student performance articles	Literature review	Positive- The performance of the students has been estimated.
Xing et al.,(2015)	Book	Compare models for predicting students' performance.	Online participating students	Testing	Positive- The model with the best results among 5 different prediction models was found.

Table 1.

Continued

Gobert et al., (2015)	Book	Evaluation of science skills.	Students	Evaluation	Positive- Student achievements increased.
Badr et al., (2016)	Article	Predicting students' probability of success in a lesson before entering the lesson.	Undergraduate students	Evaluation and grouping	Negative-Mathematics courses have no impact on their success in programming courses, it has been found.
Kim et al., (2016)	Article	Create a model for predicting students' performance.	105 Undergraduate students	Analysis, comparison	The positive-forecast model is highly accurate.
Ognjanovic et al., (2016)	Book	Estimation of student course selections.	Undergraduate students	Comparison	Positive- Course selections match predicted reality.
Ahmed et al., (2016)	Conference paper	Focuses on predicting instructor performance and improving the quality of the education system.	Instructors	Classification , comparison	Positive-Best classification algorithm found.
Hassan & Al-Razgan, (2016)	Article	Students in the Kingdom of Arabia not being able to enter any schools they want despite being successful.	High school students	Data clustering and analysis	Positive- Shows that the enrolled year affects the GPA and that the students' enrolled class does not appear in the equation.
Asif et al., (2017)	Book	Predicting their academic achievements.	Undergraduate students	Analysis	Positive-Student performances were estimated.
Costa et al., (2017)	Article	Identifying students who are likely to fail at an early stage.	Undergraduate students	Analysis	Positive- Students with the possibility of failure were identified.

Table 1.

Continued

Sandoval et al., (2018)	Article	Predicting failed students.	Course students	Data Analysis	Positive- Performs well.
Miguéis et al., (2018)	Book	To be able to determine the performance levels of the students predicted by the model.	2459 engineering students	Analysis	Positive-Best strategies are determined.
Yang & Li, (2018)	Article	Predicting student performance.	High school students	Analysis	Positive- Accurate results.
Adekitan & Salau, (2019)	Article	Students who cannot graduate are determined.	Engineering students	Analysis	Positive- 89% gave correct results.
Injadat et al., (2020)	Article	Predicting student performance.	Course students	Analysis	Positive- High accuracy.

Another important issue in the literature review is the studies that classify class the students for a purpose. About 20% of the 60 articles examined were used to divide students into groups, and these studies often used clustering methods in data mining. In an article by Natek & Zwilling (2014), they used "excel and weka" programs to compared and examined by classing student performance according to their success status in higher education. Kaur et al. (2015) classified students according to their achievements. Alfiani & Wulandari (2015) used a clustering algorithm to examine the effects of students on their achievements based on demographic characteristics such as gender and blood type. Kim et al. (2018), students who took online courses were analyzed and different models of students with an SRL profile were discussed. Amornsinlaphachai (2015), used data mining clustering and classification techniques to create a collaborative learning framework for group learners. The study of Park et al. (2016) tried to identify subtypes with the clustering approach of education data mining and the hidden class analysis method. This study was used to obtain the common activity features of 602 courses at a private university in South Korea using the "Learning Management System" and online behavior data monitored from the institution's course database. Classification and clustering procedures were applied using student data in the study.

In Table.2, the studies that classify the students for specific purposes are presented as the type, purpose, application field, method, and contribution to the literature. 12 of the 60 articles are based on research on the classification of students. When the studies are examined, 8 of the student classification studies are academic articles, 3 conference papers and 1 book type research.

Table 2.

Classifying Students

The People Who Made the Work	Type of Study	Purpose of the study / Aim (Objective)	Application Area and Sample	Metod	Conclusion- Positive/Negative- Contribution to The Field Literature
Natek & Zwilling, (2014)	Article	To classify students according to their success.	Higher education students	Comparison	Negative- Only small data can be studied.
Mayilvaganan & Kalpanadevi, (2015)	Article	To classify students according to their cognitive skills.	Students	Classification, clustering, and Analysis	Positive- Can be analyzed for students using problem-solving.
Park et al., (2016)	Article	Determining subtypes using the Latent Class Analysis method.	Undergraduate students and 602 courses	Classification and clustering	Negative- Different classification methods were not used.
Kaur et al., (2015)	Conference paper	Predicting slow learning students.	High school students' information was used.	Comparison	Positive-The best performance algorithm has been found.
Amornsinlaphach ai, (2015)	Article	Grouping learners.	Students	Analysis, examination, design, evaluation	Positive- It has been proven to be compatible.

Alfiani & Wulandari, (2015)	Conference paper	To classify students demographically.	300 students	Testing	Positive- 306 students were divided into 3 different groups.
-----------------------------	------------------	---------------------------------------	--------------	---------	--

Table 2.

Continued

Stahovich & Lin, (2016)	Book	To show the importance of classification and grouping techniques.	A database with 28 million pen strokes	Analysis	Positive- Found technique gives the most accurate result.
Rattanamethawong et al., (2018)	Article	Classifying the graduates according to their lifestyle.	300 undergraduate graduates	Analysis, grouping	Positive- It has been proven to be positive.
Priyambada et al., (2017)	Conference paper	Suggest a methodology for evaluating the curriculum according to students' behavior.	Information systems students	Clustering, testing	Positive- Possible improvement in curriculum identified.
Kim et al., (2018)	Article	To examine different models of students with SRL profiles.	284 undergraduate students enrolled in an online statistics course	Clustering	Positive- Teaching strategies that can be used to support the student are presented.
Viloria et al, (2019)	Article	To compare the classification models.	1054 article	Classification	Negative-No significant difference was found between them.
Martínez-Abad et al., (2020)	Article	Presents an innovative methodological proposal.	Data from PISA 2015	Comparison	Positive-The best classification method has been found.

The 3rd notable subject of the literature review is studies investigating the impact of technological applications in the field of data mining on students and tutorials (Özden & Atasoy, 2019). Studies to use more technology in the field of education or, in other words, to bring technology into education are listed in Table.3. The common purpose of these studies is to add technology to education. In fact, the contribution of technology to education is mentioned in all studies, but the main purpose of these 13 studies is technological developments. In the study conducted by Y. Li & Zhai, (2018), technology was planned to be used more widely in education. The research by Angeli et al.(2017) illustrates how data mining can be used in the field of educational technology to develop training program evaluation practices. Santoso, (2017) proved the advantages of a modern data warehouse instead of a traditional data warehouse. Chassignol et al. (2018) examines how technology affects students. The study of Taub et al. (2018) observed that students increase the logic of reasoning through a game. In Krau, (2015) study, it was emphasized that the use of technological tools is beneficial for students studying nursing. Patil & Kulkarni, (2018) examined how students process courses and subjects via twitter in their study.

In Table 3, studies on the research of the contribution of technological developments to education are presented as genre, purpose, application area, method, and literature. 13 of the 60 articles are based on research on student classification. When the studies are examined, 8 of the student classification studies are academic articles, 3 conference papers, and 2 book types of research.

Table 3.

The Effect of Technological Applications in the Field of Data Mining on Students and Tutors

The People Who Made the Work	Type of Study	Purpose of the study / Aim (Objective)	Application Area and Sample	Metod	Conclusion- Positive/Negative-Contribution to The Field Literature
Karal et al., (2014)	Article	To evaluate the artificial intelligence-based distance education system in terms of the contribution of students to the problem-solving process.	59 High school students	Testing	Positive- Students were satisfied with the system.
Wassan, (2015)	Article	Big data technology-based education focuses on modeling studies.	Electronic courses	Creating plans, analyzing	Positive- Suggestion is given for big data.
Krau, (2015)	Article	To investigate the effects of early warning and telehealth technologies on nurses.	Nursing students	Comparison	Positive- Technology is positive for nurses.
Patil & Kulkarni, (2018)	Article	To examine how students share their ideas on social media.	Engineering students	Comparison	Positive- Memetic algorithm worked best.
Taub et al.,(2018)	Book	Evaluating how scientific reasoning affects the completion of the game.	64 license participants	Testing and grouping	Positive- proven to influence participants based on in-game behavior.
Angeli et al., (2017)	Book	To examine the usage areas of data mining in educational technology.	One is made in Europe and the other in Australia.	Data mining and fuzzy logic	Positive- It is a definitive method for finding reliable data.

Table 3.

Continued

Ueda & Nakamura, (2017)	Conference paper	To develop data collection method.	Course of Study	Testing	Positive- A new data collection method has been developed.
Santoso, (2017)	Article	In the education system, she/he researched a modern data warehouse instead of the traditional data warehouse.	For universities	Comparison	Positive- The most suitable data warehouse has been found.
Srinivas & Rajendran, (2019)	Article	Analyzing and detecting online student comments.	Educational institutions	Analysis	Positive- Analysis was made on 12 different issues and solutions were presented
Y. Li & Zhai, (2018)	Conference paper	To make education more technological.	Technology	Literature review	Positive- There is flexibility in the education system.
Chassignol et al., (2018)	Article	To examine how technology affects behavior.	Technology and students	Literature review	Positive- The benefits of technology on students have been proven.
Viloria, López, Payares, et al., (2019)	Article	To determine the interaction of students with the distance education model (DEM).	Students	Analysis	Positive- 92.9% correct results were obtained.
Klimek & Klimek, (2020)	Conference paper	Analyzing classification trees.	239 survey	Analysis	Positive- affirms a specific principle.

Table 4.

The Others Studies

The People Who Made the Work	Type of Study	Purpose of the study / Aim (Objective)	Application Area and Sample	Method	Conclusion- Positive/Negative- Contribution to The Field Literature
Chalaris et al. (2014)	Conference paper	To support HEI management related to educational processes using data mining techniques.	For undergraduate students	Students	Positive-HEI has been proven to increase the quality of education.
Shukor et al., (2015)	Conference paper	To apply the data mining technique in which students' learning experiences can be evaluated according to their log files.	20 undergraduate students	Classification and testing	Positive- Performs well as a quiet student.
Fotache & Strimbei, (2015)	Conference paper	Today, some implications for the main coordinates of data processing and the academic curriculum are presented.	Businesses	Testing	Positive- Inference is presented.
Irfan & Gudivada, (2016)	Article	Examine the roles played by data mining and learning analytics communities in education.	Students	Examination	Positive- Progress has been made.
Burgos et al., (2018)	Book	To design a specific action plan to prevent students from leaving school.	Former student grades	Testing	Negative- Not very efficient.
Juhaňák et al., (2019)	Article	To examine students' behavior and interaction patterns within learning management systems (LMS).	In online exams	Comparison	Positive- More efficient than other studies done.

Table 4.

Continued

Zuhadar et al., (2019)	Article	Examining successful completion or failure of a STEM degree to be achieved.	College / university students	Quantitative analysis	Positive-The general inverse relationship between academic performance and attrition is shown to exist.
Valls et al. (2018)	Article	To summarize the strategic requirements for an urban design project in an architectural setup using a learning strategy.	Architecture students	Semantic and temporal data are linked	Negative- Other cities need to be examined for a definitive result.
Vieira et al., (2018)	Article	To conduct a systematic literature review in the area of analytics for visual learning.	52 conference paper	Literature review	Positive- It has helped those who will work in the field of positive-visual learning.
Rodrigues et al., (2018)	Article	Exploring the method of teaching and learning.	453 article	Literature review	Positive- Research has been done for 20 years.
Bajaj & Sharma, (2018)	Conference paper	To have a flexible solution that allows for simple and quick recognition of learning styles, it is recommended that it be put in a cloud environment.	Student and educator	Comparison	Positive- Learning in the Positive-Cloud setting has been proved to be easier.
Aljobouri et al., (2018)	Book	To recommend a new application of a rigorous approach to unsupervised learning.	2 different algorithms in chemical fields	Comparison	Positive-A new application has been made.

Table 4.

Continued

Popoola et al., (2018)	Book	Understanding and optimizing the learning environment for the realization of education.	1841 engineering undergraduate student	Comparison and analysis	Positive- Recognizing and leveraging the smart campus learning climate to achieve sustainable education.
(Aldowah et al., 2019)	Article	To provide the necessary tools that institutions can use to develop a student-oriented strategy and for continuous improvement.	402 Article	Comparison	Positive- Best result found.
Drayton-Brooks et al., (2020)	Article	The goal is to explore the use of big data and data mining from nursing practitioners' clinical training and to provide a framework for competency-based training.	Nurse	Data availability	Positive- Lots of data have been brought together and made available.

Research that is not included in the three groups mentioned above, but has different application and method areas than studies conducted using the data mining method in the field of education, is as follows. There are studies such as distance education with data mining methods, STEM, LMS, artificial intelligence, big data, and the use of data obtained from social networks. Related studies are presented in Table-4. When Table-4 is evaluated, the vast majority of applications that use data mining techniques based on student data contribute positively.

Results and Evaluation

Data mining applications are increasing all over the world due to technological advances. Data mining is widely used especially in healthcare, marketing, banking, finance, the stock market, the internet, and education. Data mining and analytics, which are commonly used in many of these sectors, also have the potential for a significant education transformation. This review study was carried out by reviewing papers in the field of data mining in the education system. 163 articles on the "Science Direct" platform for the last 6 years (2014-2020) including the terms of "education" and "data mining" was examined. However, a total of 60 studies were selected, including 35 articles, 14 conference paper, and 11 books. These 60 selected articles were collected and examined in a table in the form of the purpose of the article, the field of application, method, and its contribution to the literature.

According to the results of the analysis, studies in data mining for academic performance in education are getting more and more attention. As a result of our study, it was concluded that data mining technologies in general focus on predicting student success, classifying students according to specific purposes, and investigating their impact on students and teachers.

Also, different application areas of education such as data mining applications in education, distance education, STEM, LMS, artificial intelligence, school leaving students, wearable technologies, and virtual reality discuss. However, the research examined mainly involved studies on academic achievement.

The articles examined that those related to student performance were on sub-subjects such as success prediction, career goal, e-learning analysis, and course selection. Studies on classifying students are related to cognitive skills, learning speeds, lifestyles, and curriculum recommendations. When the research groups were analyzed in the studies, we found that studies were mostly done with university students. The studies determined that the data of undergraduate, high school, secondary school, primary school, graduate, and distance education students were used as samples.

When we examined the literature, we saw that educational data mining studies are mostly done with student data. However, we think that all stakeholders of education (students, teachers, parents, administrators, etc.) should be included in the working groups in order to obtain more accurate results in educational data mining studies.

Databases and the web environment are favored as a data collection tool, according to the study results. Today, due to the growing number of web applications, more web media is assumed to be favored as a data collection tool. The results of the analysis obtained in our study are similar to this research because the articles we examined did not include the data of parents and administrators. Achievement tests, questionnaires and attitude scales are among the other data collection tools used in research. The study conducted by Anoopkumar & Rahman, (2016) stated that in data mining analyses, classification and clustering methods are mostly used.

In future studies is recommended to investigate topics such as lifelong learning, distance education technologies, suggestion systems to increase student academic success, transferring data obtained from social networks to educational environments. We can analyze data in learning management systems and social network environments with data mining methods and achieve important learning outcomes. As a result of our study, researchers who will work in the field of education using data mining methods should pay special attention to studies that include aggregate data obtained from teachers, instructors, administrators, and parents. Also, researchers can make a groundbreaking positive contribution to the field of education by analyzing students' usage data on topics related to their lessons in social media environments with data mining. Developing data mining applications with a user-friendly interface without the need for high-level programming knowledge in the field of education is also considered to be useful.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Çıkar Çatışması ve Etik Bildirimi

Yazarlar, aralarında çıkar çatışması bulunmadığını ve çalışmaya eşit oranda katkı sunduklarını beyan etmiştir. Yazarlar, tüm etik kurallara uyduklarını bildirmiştir.

References

- Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, 5(2). <https://doi.org/10.1186/s13677-020-00165-y>
- Agarwal, S., Pandey, G. N., & Tiwari, M. D. (2012). Data mining in education: data classification and decision tree approach. *International Journal of E-Education, e-Business, e-Management and e-Learning*, 2(2), 140.
- Ahmed, A. M., Rizaner, A., & Ulusoy, A. H. (2016). Using data mining to predict instructor performance. *Procedia Computer Science*, 102, 137-142. <https://doi.org/10.1016/j.procs.2016.09.380>

- 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. <https://doi.org/10.1016/j.tele.2019.01.007>
- Alfiani, A. P., & Wulandari, F. A. (2015). Mapping student's performance based on data mining approach (a case study). *Agriculture and Agricultural Science Procedia*, 3, 173–177. <https://doi.org/10.1016/j.aaspro.2015.01.034>
- Aljoubouri, H. K., Jaber, H. A., Kocak, O. M., Algin, O., & Cankaya, I. (2018). Clustering fMRI data with a robust unsupervised learning algorithm for neuroscience data mining. *Journal of Neuroscience Methods*, 299, 45–54. <https://doi.org/10.1016/j.jneumeth.2018.02.007>
- Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on big data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, 24(1), 1–7. <https://doi.org/10.1016/j.iedeen.2017.06.002>
- Amornsinlaphachai, P. (2015). The design of a framework for cooperative learning through web utilizing data mining technique to group learners. *Procedia-Social and Behavioral Sciences*, 174, 27–33. <https://doi.org/10.1016/j.sbspro.2015.01.622>
- Angeli, C., Howard, S. K., Ma, J., Yang, J., & Kirschner, P. A. (2017). Data mining in educational technology classroom research: Can it make a contribution? *Computers & Education*, 113, 226–242. <https://doi.org/10.1016/j.compedu.2017.05.021>
- Anoopkumar, M., & Rahman, A. M. J. M. Z. (2016). A Review on data mining techniques and factors used in educational data mining to predict student amelioration. *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 122–133. <https://doi.org/10.1109/SAPIENCE.2016.7684113>
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Aydoğdu, Ş. (2020). Educational data mining studies in Turkey: A systematic review. *Turkish Online Journal of Distance Education*, 21(3), 170–185. <https://doi.org/10.17718/tojde.762046>
- Badr, G., Algobail, A., Almutairi, H., & Almutery, M. (2016). Predicting students' performance in university courses: a case study and tool in KSU mathematics department. *Procedia Computer Science*, 82, 80–89. <https://doi.org/10.1016/j.procs.2016.04.012>
- Bajaj, R., & Sharma, V. (2018). Smart education with artificial intelligence based determination of learning styles. *Procedia Computer Science*, 132, 834–842. <https://doi.org/10.1016/j.procs.2018.05.095>
- Baker, R. (2010). Data mining for education. *International Encyclopedia of Education*, 7(3),

112–118.

- Balaman, S. (2020). A study on the impacts of digital storytelling on EFL learners' self-efficacy and attitudes toward education technologies. *International Online Journal of Education and Teaching*, 7(1), 289–311.
- Bhullar, M. S., & Kaur, A. (2012). Use of data mining in education sector. *Proceedings of the World Congress on Engineering and Computer Science*, 1, 24–26.
- Burgos, C., Campanario, M. L., de la Peña, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers & Electrical Engineering*, 66, 541–556. <https://doi.org/10.1016/j.compeleceng.2017.03.005>
- Cabada, R. Z., Estrada, M. L. B., & Bustillos, R. O. (2018). Mining of educational opinions with deep learning. *Journal of Universal Computer Science*, 24(11), 1604–1626.
- Campagni, R., Merlini, D., Sprugnoli, R., & Verri, M. C. (2015). Data mining models for student careers. *Expert Systems with Applications*, 42(13), 5508–5521. <https://doi.org/10.1016/j.eswa.2015.02.052>
- Chakraborty, B., Chakma, K., & Mukherjee, A. (2016). A density-based clustering algorithm and experiments on student dataset with noises using rough set theory. *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, 431–436. <https://doi.org/10.1109/ICETECH.2016.7569290>
- Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C., & Tsolakidis, A. (2014). Improving quality of educational processes providing new knowledge using data mining techniques. *Procedia-Social and Behavioral Sciences*, 147, 390–397. <https://doi.org/10.1016/j.sbspro.2014.07.117>
- Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia Computer Science*, 136, 16–24. <https://doi.org/10.1016/j.procs.2018.08.233>
- Chen, J., Wei, W., Guo, C., Tang, L., & Sun, L. (2017). Textual analysis and visualization of research trends in data mining for electronic health records. *Health Policy and Technology*, 6(4), 389–400.
- Converse, G., Curi, M., & Oliveira, S. (2019). Autoencoders for educational assessment. *International Conference on Artificial Intelligence in Education*, 41–45. https://doi.org/10.1007/978-3-030-23207-8_8
- Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73, 247–256. <https://doi.org/10.1016/j.chb.2017.01.047>

- Drayton-Brooks, S. M., Gray, P. A., Turner, N. P., & Newland, J. A. (2020). The use of big data and data mining in nurse practitioner clinical education. *Journal of Professional Nursing*. <https://doi.org/10.1016/j.profnurs.2020.03.012>
- Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, *94*, 335–343. <https://doi.org/10.1016/j.jbusres.2018.02.012>
- Fok, W. W. T., He, Y. S., Yeung, H. H. A., Law, K. Y., Cheung, K. H., Ai, Y. Y., & Ho, P. (2018). Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine. *2018 4th International Conference on Information Management (ICIM)*, 103–106. <https://doi.org/10.1109/INFOMAN.2018.8392818>
- Fotache, M., & Strimbei, C. (2015). SQL and data analysis. Some implications for data analysis and higher education. *Procedia Economics and Finance*, *20*, 243–251. [https://doi.org/10.1016/S2212-5671\(15\)00071-4](https://doi.org/10.1016/S2212-5671(15)00071-4)
- Gobert, J. D., Kim, Y. J., Sao Pedro, M. A., Kennedy, M., & Betts, C. G. (2015). Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld. *Thinking Skills and Creativity*, *18*, 81–90.
- Guan, X., Fan, Y., Qin, Q., Deng, K., & Yang, G. (2020). Construction of science and technology achievement transfer and transformation platform based on deep learning and data mining technology. *Journal of Intelligent & Fuzzy Systems, Preprint*, 1–12.
- Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting students performance in educational data mining. *2015 International Symposium on Educational Technology (ISET)*, 125–128. <https://doi.org/10.1109/ISET.2015.33>
- Guo, J., Gu, C., Yang, J., Zhang, Y., & Yang, H. (2020). Data mining and application of ship impact spectrum acceleration based on PNN neural network. *Ocean Engineering*, *203*, 107193. <https://doi.org/10.1016/j.oceaneng.2020.107193>
- Gupta, P., Mehrotra, D., & Sharma, T. K. (2015). Identifying knowledge indicators in higher education organization. *Procedia Computer Science*, *46*, 449–456. <https://doi.org/10.1016/j.procs.2015.02.043>
- Hassan, S. M., & Al-Razgan, M. S. (2016). Pre-university exams effect on students GPA: a case study in IT department. *Procedia Computer Science*, *82*, 127–131. <https://doi.org/10.1016/j.procs.2016.04.018>
- Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A systematic review of deep learning approaches to educational data mining. *Complexity*, *2019*. <https://doi.org/10.1155/2019/1306039>

- Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.-X., & Xu, C. (2018). Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. *Science of the Total Environment*, 630, 1044–1056. <https://doi.org/10.1016/j.scitotenv.2018.02.278>
- Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. *Knowledge-Based Systems*, 105992. <https://doi.org/10.1016/j.knosys.2020.105992>
- Irfan, M. T., & Gudivada, V. N. (2016). Cognitive computing applications in education and learning. In *Handbook of Statistics* (Vol. 35, pp. 283–300). Elsevier. <https://doi.org/10.1016/bs.host.2016.07.008>
- Jiang, Y., & Li, X. (2020). Intelligent online education system based on speech recognition with specialized analysis on quality of service. *International Journal of Speech Technology*, 1–9. <https://doi.org/10.1007/s10772-020-09723-w>
- Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, 92, 496–506. <https://doi.org/10.1016/j.chb.2017.12.015>
- Karal, H., Nabiyeve, V., Erümit, A. K., Arslan, S., & Çebi, A. (2014). Students' opinions on artificial intelligence based distance education system (Artimat). *Procedia-Social and Behavioral Sciences*, 136, 549–553. <https://doi.org/10.1016/j.sbspro.2014.05.374>
- 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. <https://doi.org/10.1016/j.procs.2015.07.372>
- Kim, D., Park, Y., Yoon, M., & Jo, I.-H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *The Internet and Higher Education*, 30, 30–43. <https://doi.org/10.1016/j.iheduc.2016.03.002>
- Kim, D., Yoon, M., Jo, I.-H., & Branch, R. M. (2018). Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea. *Computers & Education*, 127, 233–251. <https://doi.org/10.1016/j.compedu.2018.08.023>
- Klimek, J., & Klimek, J. A. (2020). IT and data mining in decision-making in the organization. Education management in the culture of late modernity. *Procedia Computer Science*, 176, 1990–1999. <https://doi.org/10.1016/j.procs.2020.09.235>
- Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rose, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333–353. <https://doi.org/10.1002/wcs.1350>
- Krau, S. D. (2015). The influence of technology in nursing education. *Nursing Clinics*, 50(2),

- 379–387. <https://doi.org/10.1016/j.cnur.2015.02.002>
- Lang, S., Bravo-Marquez, F., Beckham, C., Hall, M., & Frank, E. (2019). Wekadeeplearning4j: A deep learning package for weka based on deeplearning4j. *Knowledge-Based Systems*, 178, 48–50. <https://doi.org/10.1016/j.knosys.2019.04.013>
- Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., & Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European higher education area—application to student data from Open University of Madrid, UDIMA. *Computers & Education*, 72, 23–36. <https://doi.org/10.1016/j.compedu.2013.10.009>
- Li, X., & Wu, X. (2015). Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4520–4524. <https://doi.org/10.1109/ICASSP.2015.7178826>
- Li, Y., & Zhai, X. (2018). Review and prospect of modern education using big data. *Procedia Computer Science*, 129, 341–347. <https://doi.org/10.1016/j.procs.2018.03.085>
- Liao, S. (2003). Knowledge management technologies and applications—literature review from 1995 to 2002. *Expert Systems with Applications*, 25(2), 155–164. [https://doi.org/10.1016/S0957-4174\(03\)00043-5](https://doi.org/10.1016/S0957-4174(03)00043-5)
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M.-J. (2020). Educational data mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, 66, 100875. <https://doi.org/10.1016/j.stueduc.2020.100875>
- Mayilvaganan, M., & Kalpanadevi, D. (2015). Cognitive skill analysis for students through problem solving based on data mining techniques. *Procedia Computer Science*, 47, 62–75. <https://doi.org/10.1016/j.procs.2015.03.184>
- Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115, 36–51. <https://doi.org/10.1016/j.dss.2018.09.001>
- Natek, S., & Zwilling, M. (2014). Student data mining solution—knowledge management system related to higher education institutions. *Expert Systems with Applications*, 41(14), 6400–6407. <https://doi.org/10.1016/j.eswa.2014.04.024>
- Notten, T. (2014). Public management and leadership. Lasagne model in training for educational management. *Journal of Social Intervention: Theory and Practice*, 23(1), 144–148.
- Ognjanovic, I., Gasevic, D., & Dawson, S. (2016). Using institutional data to predict student course selections in higher education. *The Internet and Higher Education*, 29, 49–62. <https://doi.org/10.1016/j.iheduc.2015.12.002>
- Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for

- students' performance prediction. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 598–599. <https://doi.org/10.1145/3027385.3029479>
- Özden, C., & Atasoy, R. (2019). Determination of educational needs of technology and design courses in secondary school students. *International Online Journal of Education and Teaching*, 6(3), 511–523.
- Pandey, U. K., & Bhardwaj, B. K. (2012). Data Mining as a torch bearer in education sector. *ArXiv Preprint ArXiv:1201.5182*.
- Park, Y., Yu, J. H., & Jo, I.-H. (2016). Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education*, 29, 1–11. <https://doi.org/10.1016/j.iheduc.2015.11.001>
- Patil, S., & Kulkarni, S. (2018). Mining social media data for understanding students' learning experiences using memetic algorithm. *Materials Today: Proceedings*, 5(1), 693–699. <https://doi.org/10.1016/j.matpr.2017.11.135>
- Popoola, S. I., Atayero, A. A., Badejo, J. A., John, T. M., Odukoya, J. A., & Omole, D. O. (2018). Learning analytics for smart campus: data on academic performances of engineering undergraduates in Nigerian private university. *Data in Brief*, 17, 76–94. <https://doi.org/10.1016/j.dib.2017.12.059>
- Priyambada, S. A., Mahendrawathi, E. R., & Yahya, B. N. (2017). Curriculum assessment of higher educational institution using aggregate profile clustering. *Procedia Computer Science*, 124, 264–273. <https://doi.org/10.1016/j.procs.2017.12.155>
- Rattanamethawong, N., Sinthupinyo, S., & Chandrachai, A. (2018). An innovation model of alumni relationship management: Alumni segmentation analysis. *Kasetsart Journal of Social Sciences*, 39(1), 150–160. <https://doi.org/10.1016/j.kjss.2017.02.002>
- Reamer, A. C., Ivy, J. S., Vila-Parrish, A. R., & Young, R. E. (2015). Understanding the evolution of mathematics performance in primary education and the implications for STEM learning: A Markovian approach. *Computers in Human Behavior*, 47, 4–17. <https://doi.org/10.1016/j.chb.2014.09.037>
- Rodrigues, M. W., Isotani, S., & Zárate, L. E. (2018). Educational data mining: A review of evaluation process in the e-learning. *Telematics and Informatics*, 35(6), 1701–1717. <https://doi.org/10.1016/j.tele.2018.04.015>
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146. <https://doi.org/10.1016/j.eswa.2006.04.005>
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27. <https://doi.org/10.1002/widm.1075>

- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- Saleti, S., & Subramanyam, R. B. V. (2019). A MapReduce solution for incremental mining of sequential patterns from big data. *Expert Systems With Applications*, 133, 109–125. <https://doi.org/10.1016/j.eswa.2019.05.013>
- Sandoval, A., Gonzalez, C., Alarcon, R., Pichara, K., & Montenegro, M. (2018). Centralized student performance prediction in large courses based on low-cost variables in an institutional context. *The Internet and Higher Education*, 37, 76–89. <https://doi.org/10.1016/j.iheduc.2018.02.002>
- Santoso, L. W. (2017). Data warehouse with big data technology for higher education. *Procedia Computer Science*, 124, 93–99. <https://doi.org/10.1016/j.procs.2017.12.134>
- Sen, B., & Ucar, E. (2012). Evaluating the achievements of computer engineering department of distance education students with data mining methods. *Procedia Technology*, 1, 262–267. <https://doi.org/10.1016/j.protcy.2012.02.053>
- Shabtay, L., Fournier-Viger, P., Yaari, R., & Dattner, I. (2020). A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. *Information Sciences*. <https://doi.org/10.1016/j.ins.2020.10.020>
- Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414–422. <https://doi.org/10.1016/j.procs.2015.12.157>
- Shukor, N. B. A., Tasir, Z., & van der Meijden, H. A. T. (2015). *An examination of online learning effectiveness using data mining*. <https://doi.org/10.1016/j.sbspro.2015.01.402>
- Srinivas, S., & Rajendran, S. (2019). Topic-based knowledge mining of online student reviews for strategic planning in universities. *Computers & Industrial Engineering*, 128, 974–984. <https://doi.org/10.1016/j.cie.2018.06.034>
- Stahovich, T. F., & Lin, H. (2016). Enabling data mining of handwritten coursework. *Computers & Graphics*, 57, 31–45. <https://doi.org/10.1016/j.cag.2016.01.002>
- Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and Instruction*, 54, 93–103. <https://doi.org/10.1016/j.learninstruc.2017.08.005>
- Tayefi, M., Tajfard, M., Saffar, S., Hanachi, P., Amirabadizadeh, A. R., Esmaeily, H., Taghipour, A., Ferns, G. A., Moohebaty, M., & Ghayour-Mobarhan, M. (2017). hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. *Computer Methods and Programs in Biomedicine*, 141, 105–109. <https://doi.org/10.1016/j.cmpb.2017.02.001>

- Thuraisingham, B. (2003). *Web data mining and applications in business intelligence and counter-terrorism*. CRC Press.
- Ueda, H., & Nakamura, M. (2017). Data analysis for evaluation on course design and improvement of “cyberethics” moodle online courses. *Procedia Computer Science*, 112, 2345–2353. <https://doi.org/10.1016/j.procs.2017.08.204>
- Valls, F., Redondo, E., Fonseca, D., Torres-Kompen, R., Villagrasa, S., & Martí, N. (2018). Urban data and urban design: A data mining approach to architecture education. *Telematics and Informatics*, 35(4), 1039–1052. <https://doi.org/10.1016/j.tele.2017.09.015>
- Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, 122, 119–135. <https://doi.org/10.1016/j.compedu.2018.03.018>
- Viloria, A., López, J. R., Leyva, D. M. G., Vargas-Mercado, C., Hernández-Palma, H., Llinas, N. O., David, M. A., & Rodríguez, J. V. (2019). Data mining techniques and multivariate analysis to discover patterns in university final researches. *Procedia Computer Science*, 155, 581–586. <https://doi.org/10.1016/j.procs.2019.08.081>
- Viloria, A., López, J. R., Payares, K., Vargas-Mercado, C., Duran, S. E., Hernández-Palma, H., & David, M. A. (2019). Determinating student interactions in a virtual learning environment using data mining. *Procedia Computer Science*, 155, 587–592. <https://doi.org/10.1016/j.procs.2019.08.082>
- Vranic, M., Pintar, D., & Skocir, Z. (2007). The use of data mining in education environment. *2007 9th International Conference on Telecommunications*, 243–250. <https://doi.org/10.1109/CONTEL.2007.381878>
- Wang, H., Tan, X., Huang, Z., Pan, B., & Tian, J. (2020). Mining incomplete clinical data for the early assessment of Kawasaki disease based on feature clustering and convolutional neural networks. *Artificial Intelligence in Medicine*, 101859. <https://doi.org/10.1016/j.artmed.2020.101859>
- Wang, L., Sy, A., Liu, L., & Piech, C. (2017). Learning to represent student knowledge on programming exercises using deep learning. *International Educational Data Mining Society*.
- Wassan, J. T. (2015). Discovering big data modelling for educational world. *Procedia-Social and Behavioral Sciences*, 176, 642–649. <https://doi.org/10.1016/j.sbspro.2015.01.522>
- Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, 47, 168–181. <https://doi.org/10.1016/j.chb.2014.09.034>
- Xu, X., Yin, X., & Chen, X. (2019). A large-group emergency risk decision method based on

- data mining of public attribute preferences. *Knowledge-Based Systems*, 163, 495–509. <https://doi.org/10.1016/j.knsys.2018.09.010>
- 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. <https://doi.org/10.1016/j.compedu.2018.04.006>
- Yucel, M., Aslan, Z., & Burunkaya, M. (2020). Classification of the temperature-dependent gain of an erbium-doped fiber amplifier by using data mining methods. *Optik*, 164515. <https://doi.org/10.1016/j.ijleo.2020.164515>
- Zhang, H., Huang, T., Liu, S., Yin, H., Li, J., Yang, H., & Xia, Y. (2020). A learning style classification approach based on deep belief network for large-scale online education. *Journal of Cloud Computing*, 9, 1–17. <https://doi.org/10.1186/s13677-020-00165-y>
- Zhao, S., Su, Z., & Miao, G. (2020). Application of English education information management system based on convolution neural network classification algorithm. *The International Journal of Electrical Engineering & Education*. <https://doi.org/10.1177/0020720920940614>
- Zhuhadar, L., Daday, J., Marklin, S., Kessler, B., & Helbig, T. (2019). Using survival analysis to discovering pathways to success in mathematics. *Computers in Human Behavior*, 92, 487–495. <https://doi.org/10.1016/j.chb.2017.12.016>

Genişletilmiş Özet

Teknolojideki hızlı ilerlemeler ve bulut sistemlerin yaygınlaşarak eğitim alanlarında kullanıma açık duruma gelmesiyle bu alandaki verilerin anlamlı hale getirilmesi eğitimi daha etkili ve verimli hale getirebilecektir. Eğitim alanındaki işlenmemiş verilerin anlamlı hale getirilmesi sayesinde, bu alanda çalışan öğretmenlere olumlu yönde katkı sağlanması ve öğrencileriyle ilgili destekleyici bilgileri edinmeleri amacıyla veri madenciliğinin kullanılması önem arz etmektedir. Bilgisayar sistemlerinin her geçen gün maliyetlerinin düşmesi ve performanslarının artıyor olması nedeniyle, bilgisayar sistemlerinde yapılan işlemlerin artması ve daha büyük miktarda verinin oluşmasına ve dolayısıyla saklayabilme ihtiyacına ve imkanına gerek duyulmaktadır. Bu nedenle ham şekilde saklanan verilerin çok büyük boyutlara ulaştığı görülmektedir. Bu ham veriyi bilgiye veya anlamlı hale dönüştürme işlemleri veri madenciliği ile yapılabilmektedir. Veri madenciliği işlemlerinin uygulama alanı oldukça geniştir. Bu alanlar içerisinde veri madenciliğinin kullanıldığı bazı uygulama alanları şunlardır: Sağlık, endüstri, kamu, pazar araştırması, bankacılık, finans ve borsa, internet, risk analizi, kurum kaynaklarının kullanımı, ticaret, müşteri analizi, alışveriş, bankacılık, telekomünikasyon, tıp, bilim, mühendislik, sigortacılık ve eğitim gibi alanlardır. Teknolojideki hızlı gelişmelerin, öğretim teknolojileri kavramı altında eğitim alanına yansdığı görülmektedir. Veri madenciliği alanında eğitim alanına yansıyan ve kullanım alanı gün geçtikçe yaygınlaşan önemli gelişmelerdendir. Genel

olarak, veri madenciliği uygulamalarının oluşturduğu bilgi, eğitim öğretim sürecinin değerlendirilmesi, izlenmesi ve kişiselleştirilmesi gibi eğitim yöntemlerinin geliştirilmesini objektif olarak amaçlamaktadır. Örneğin, eğitimciler, öğrencilerin öğrenme düzeylerini ve akademik başarılarını izleyebilir ve eğitsel periyodu yeniden uyarlayabilirler. Bu sayede öğrencilerin ihtiyacına göre yaklaşımlarını ve öğretim etkinliklerini değiştirebilirler. Ayrıca öğrenciler kendi öz-değerlendirme süreçlerini veri madenciliği uygulamaları sayesinde yapabilirler.

Yapılan çalışmalar incelendiğinde, eğitim alanındaki verilerin anlamlı hale getirilmesi ile eğitimin daha etkili ve verimli hale getirilmesi veri madenciliği yöntemleri ile mümkündür. Veri madenciliği destekli eğitim çalışmaları incelendiğinde e-öğrenme sistemlerinin geliştirmesi, pedagojik destek, eğitim verilerinin kümelenmesi, öğrenci performans tahminleri gibi konuları içermektedir. Ancak, bulut sistemlerinin eğitim alanında kullanılması ve makine öğrenmesi tekniklerinin veri madenciliği yöntemlerinde özellikle son yıllarda çok daha kapsamlı ve geniş imkanlar sunmaktadır.

Bu çalışmada veri madenciliği yöntemlerinin eğitim alanında hangi yöntemlerle ve hangi uygulama alanlarında nasıl kullandıklarının tespit edilmesine dair bir literatür taraması yapılmıştır. Araştırmada; veri madenciliği yöntemi kullanılarak eğitim alanında yapılan çalışmaları belirli başlıklar altında gruplamak, yöntemlerini ve amaçlarını belirlemek ve bu alanda çalışacak olan kişilere alandaki ihtiyacı sunmak ve yön göstermek amaçlanmıştır. Bu sayede daha sonra veri madenciliği yöntemlerini kullanarak eğitim alanında uygulama yapmak isteyen kişilere katkı sağlayacağı öngörülmektedir. Ayrıca güncel uygulama alanlarını, yöntemleri ve çalışma amaçlarını sunarak literatürde ihtiyaç duyulan konuların belirlenmesi amaçlanmıştır. Araştırmamızda eğitim alanında veri madenciliği yöntemi kullanılarak 2014 ile 2020 yılları arasında yapılmış araştırmalar “Science Direct” platformu üzerinden taranmıştır. Veri Madenciliği alanında yapılan çalışmalar özellikle “Hangi spesifik alanda ve ne şekilde uygulandı?” sorusuna cevap verecek şekilde incelenmiştir. Araştırma sonucunda 163 makale içerisinden hedefe uygun olan ve doğrudan eğitimde veri madenciliği ile ilgili 60 makale seçilmiştir. Makalelerin içerik analizi yöntemiyle gruplama ve tablo gösterimi ile sunulmuştur. Seçilen 35 adet makale, 14 adet konferans bildirisi, 11 adet kitap olmak üzere toplamda 60 adet çalışma, çalışmanın amacı, uygulama alanı ve örneklem, metot ve yöntemi, literatüre katkısı şeklinde tasnif edilerek sunulmuştur. Sunulan makaleler çalışmanın amacına göre öğrenci başarılarının tahmin edilmesi, öğrencilerin belirli amaçlar için sınıflandırılması, veri madenciliği alanındaki teknolojik uygulamaların öğrenciler ve öğretmenler üzerine etkisini araştırılması ve diğerleri şeklinde dört temel tabloda incelenmiştir. Sunulan tablolar birbirinden tamamen ayrı kümeler halinde olmayıp, çalışmanın esas amacı göz önüne alınarak gruplandırılmıştır.

Günümüzde gelişen veri madenciliği teknolojileri ile birlikte öğrenci başarılarının tahmin edilmesi, öğrencilerin belirli amaçlar için sınıflandırılması, veri madenciliği

alanındaki teknolojik uygulamaların öğrenciler ve öğretmenler üzerine etkisini araştırılması ve uzaktan eğitim, STEM, LMS, yapay zeka gibi eğitimin farklı uygulama alanlarını ele alındığı analiz sonucunda tespit edilmiştir. Ancak incelenen araştırmalarda ağırlıklı olarak akademik başarı üzerine yapılan çalışmaların yer aldığı görülmüştür. İncelenen araştırmalarda öğrenci performansı ile alakalı olanların başarı tahmini, kariyer hedefi, e-öğrenme analizi, ders seçimi gibi konularında olduğu görülmüştür. Öğrencileri sınıflandırma ile ilgili çalışmaların ise bilişsel beceriler, öğrenme hızları, yaşam tarzları ve müfredat önerileri gibi konularda olduğu görülmüştür. Teknolojik uygulamaların öğrenciler ve öğretmenler üzerine etkilerini araştıran çalışmalarda teknolojinin öğrenme sürecini hızlandıran faktörler üzerinde değerlendirildiği görülmüştür. Bu üç grupta yer almayan çalışmaların ise öğrenme süreci, okuldan ayrılan öğrenciler, LMS, STEM vb. konuların araştırıldığı tespit edilmiştir. Çalışma grubuna ilişkin veriler incelendiğinde bu çalışmada çoğunlukla üniversite öğrencileri ile çalışmaların yapıldığı görülmektedir. Lise, ortaokul, ilköğretim, yüksek lisans ve uzaktan eğitim ile kurslara kayıtlı öğrenciler hakkında da veriler de kullanıldığı tespit edilmiş fakat çalışmalarda daha az yer aldığı sonucuna ulaşılmıştır.

Veri madenciliği ile ilgili yapılacak çalışmalarda da çalışma gruplarının eğitimle ilgilenen tüm kişileri (öğrenciler, öğretmenler, veliler, yöneticiler vb.) kapsayacak şekilde gerçekleştirilmesi gerektiği düşünülmektedir. Analiz sonucuna göre, veri toplama aracı olarak veritabanları ve web ortamının tercih edildiği görülmektedir. Günümüzde web uygulamalarının artması nedeniyle veri toplama aracı olarak, daha çok web ortamının tercih edildiği düşünülmektedir. Başarı testleri, anket ve tutum ölçekleri de araştırmalarda kullanılan diğer veri toplama araçlarından olmuştur. Yapılan bu çalışma ile veri madenciliği yöntemi kullanılarak eğitim alanında çalışma yapacak araştırmacılara katkı sağlanacağı düşünülmektedir. Öğrencilerin, öğretmenlerin, yöneticilerin ve velilerin verilerinin veri madenciliği yöntemleri ile anlamlı hale getiren uygulama içeren çalışmaların alana katkı sağlayabileceği düşünülmektedir. Ayrıca eğitim alanında programlama bilgisi ve teknik bilgi gereksizdir kullanıcı dostu arayüze sahip veri madenciliği araçlarının üretilmesinin yararlı olabileceği düşünülmektedir.