DATA MINING AND UNIVERSITY DROPOUT: A LITERATURE REVIEW

HENRY DAVID BACCA MORALES¹ EDWIN HERNANDO ZAPATA CIFUENTES² &
CARLOS ENRIQUE MONTENEGRO MARÍN³

¹ Facultad de Ingeniería, Diseño e Innovación / Institución Universitaria Politécnico Grancolombiano

² Facultad de Ingeniería, Diseño e Innovación / Institución Universitaria Politécnico Grancolombiano

³ Facultad de Ingeniería / Universidad Distrital Francisco José de

ABSTRACT: Nowadays, data mining has been applied to the educational field as a path for problem solving such as school dropout. There are different investigations that highlight the importance of the processing of large amounts of data to predict and prevent the desertion of students by the implications that this brings at the social and economic level. This study analyzes some key concepts and literature available on data mining and school dropout showing the current state of these thematic, specifically in the IEEE XPLORE Digital Library online database which is a collection of highly recognized research at a global level.

KEY WORDS: Data Mining, School Dropout, Educational Data Mining & Literature Review

Received: Apr 08, 2021; **Accepted:** Apr 28, 2021; **Published:** May 08, 2021; **Paper Id.:** IJMPERDJUN202125

1. INTRODUCTION

Educational data mining "is a research field that seeks to extract useful information from large sets of educational data. This area provides tools to improve student retention rates around the world "[1], the analysis and interventions concerning their performance or desertion and thus be able to take timely actions in order to solve problems and needs that affect the educational sector and his actors.

University desertion is a current concern in most institutions around the world, producing low rate of enrollment, decrease in assets and economic difficulties at the state level, and finally different social obstacles in general [2]. According to Pereira, RT and Zambrano, JC (2020) "one of the main problems facing the Colombian Higher Education System has to do with high levels of desertion" [2].

According to the Ministry of Education in Colombia (2009) in recent years, a considerable increase in the number of students enrolled, however, only a small part of them manages to end their studies and graduate from them. This would indicate that there is a high level of abandonment especially in the first semesters, according to this, "of each hundred students who get into a university approximately half cannot complete their academic year and obtain graduation" [3].

On the other hand, school desertion is a problem that currently prejudices educational quality; There are several investigations that coincide in defining some variables linked to graduation, some of them are related to differences in the location, the demographics of students and financing. Some studies mention that there is not a specific reason that explains why students abandon school, "in fact, it is a multifactorial problem called the problem of the thousand factors " [4].

Therefore, it is affirmed that data mining may facilitate early detection of possible deserters students, since, "if the factors that contribute significantly to abandonment through statistical inferences can be identified, then it is useful for developing strategies for preventing the abandonment" [5], this will allow the institution to act early and consequently, that the student does not prematurely abandon its training process in the institution by generating, for example, study plans based on needs, tutoring or consulting for the areas that require it; this before the desertion process is given and even previously the student starts his studies.

The thrust topics of this article have been worked by different authors worldwide and for several years, so this study seeks to explore some of these literary works and show the current state of these thematic, specifically at the data online base IEEE XPLORE Digital Library which is a collection highly and globally recognized.

According to the above, some key concepts will be defined which are fundamental and common to the current study, initiating with the desertion that it is one of the most important criteria of this article. Then an analysis of the literary review will be done, the result of this and finally the conclusions will be given according to the literature review.

2. KEY CONCEPTS

School Dropout: For this paper the taken concept about school dropout is, according to Páramo and Correa (1999), "the school dropout, understood not only as the definitive abandonment of classrooms, but as the abandonment of academic training" [6]. The desertion of students can be presented in any modality of education, taking into account not only the face-to-face education services but also those that correspond to e-learning.

School dropout as mentioned above has different causes, therefore it is important that a methodology that allows them to establish them and also predict this abandonment in advance. According to Stefanova and Kabakchieva (2017), data mining is an option that has been recently used in the educational field since large data processing is usually used in business sectors, but there are few cases where it applies to the educational sector [7]. The Big Data refers to the large amounts of data that cannot be processed by a common software, hence there are data mining techniques that allow classification and organize the information for benefit, in this case, of the prediction of school dropout; for this reason, it is necessary to define the notion of data mining, this other fundamental concept being for this document.

Data Mining: According to Zhang and Qin (2018) data mining is a process of extraction of large amounts of data, some of them inconclusive, diffuse and selected, in which there is confusing information and knowledge, but at the end, they are probably useful [8]. In this way data mining is used in fields such as engineering, medicine, economics and education.

a. Data Mining Techniques

Data mining techniques allow to systematically examine large amounts of data to demonstrate their behavior in a defined environment. Thus, there are different data mining techniques, among the most relevant can be found:

Artificial neural networks: It is a set of processing networks that are interconnected and feeding on the information provided by it. Similar to neuronal networks of living beings, it is a model used for resolution of complex problems where a large number of variables that interact among them.

Nearest K-Neighbors: Rojanavas (2019) defines it as "an easy, simple and non-parametric learning. The KNN algorithm is generally used to segregate the data according to the similarities in the metrics. In this algorithm, normal data points occur

around a thick neighborhood and is called a memory-based technique because the data points must be stored in the memory during the runtime. " [9].

Grouping or Clustering: For Sin and Muthu (2015) involves grouping of records that are similar, identifying the distance between them in a space of n dimensions, where n is the number of variables [10].

Classification: According to Sin and Muthu (2015) it is the identification of the category or class to which a value belongs, based on values that have been previously categorized [10].

Regression: It is a mathematical procedure to discover the connection between the dependent and independent variables. "The regression analysis is divided into linear regression, multiple regression and non-linear regression. In linear regression, the data is modeled using a straight line; the multivariate regression is an extension of linear regression, which involves multiple variables; non-linear regression is a non-linear model with the addition of a polynomial form in the basic linear model" [8].

b. Algorithms used in Data Mining

Algorithms are a set of rules that are used to solve problems taking into account concrete indications; For data mining techniques there are different algorithms, hereunder the most important are related:

Decision Trees: According to Pereira and Zambrano (2017) "a decision tree is a tree structure similar to a flowchart, where each internal node denotes a test in an attribute, each branch represents an outcome of the test and the leaf nodes represent classes. The upper node of a tree is the root node. It is a prediction model used from a database that manufactures logical construction diagrams" [11].

Random forest: According to Murakami "is an automatic learning algorithm that uses set learning. This algorithm uses multiple decision trees such as weak classifiers and integrates the results (it means, the forest) to obtain correct results. The random forest function is available to use pattern recognition, regression and cluster pool" [12].

Naive Bayes: According to Zheng, Gao, Wang and Fu (2020) "is a quasi-linear classification algorithm of rhythm in the supervised learning method. The naive model of Bayes calculates the frequency of each category in the data of the training assembly and the conditional probability of each divided category according to each characteristic value, simplifies the modeling of a joint probability distribution to model each own value. " [13].

Classification of gradient increase: "is a set method composed of an initial estimator followed by a sequence of regression decision trees. This sequence of regression decision trees is called steps in the training process of the gradient increase classifier. In each step, a single decision tree returneur is trained on the residue of the prediction scores of the step before the fundamental truth. The final prediction score is a weighted sum of the initial score and the predicted residues for each regression decision tree in the sequence" [14].

Support vector machine: SVM "is a learning method commonly used for binary classification applications. The focus of the support vector machine is based on the VC dimension of the theory of statistical learning and the minimum risk principle of St. structure. The support vector machine model can solve the problem of recognition of high-dimensional, non-linear and small sample patterns. It imitates the non-linearity of the space of entry to the high-dimension features, which is different from the method of reducing dimensionality commonly used in the recognition of traditional patterns, which allows the linear not classifiable in the low-dimension space separate linearly in the high-dimension" [13].

3. LITERATURE REVIEW

As mentioned above, the topics of data mining applied in education and school dropout have been worked considerably by different authors, in this way a broad revision of material that will encompass not only the enunciated concepts but the published productions in a period of four years (from 2017 to 2020).

Reviewed literary products were obtained from the IEEE XPLORE Digital Library online database where for each of the topics of interest, both conferences, such as articles and magazines were found. Below is the results ratio for each search:

Figure 1: Publications per Year.

According to Figure 1, it is possible to evidence that publications about school dropout had been increased from 2017 to 2019, however in 2020, it was presented to a lesser extent, (nevertheless, it is worth noting that the situation caused by the pandemic by Covid-19 was able to influence the number of publications carried out), even so it was slightly lower than the productions of 2018; In relation to publications on educational data mining these are the most significant in terms of the number of productions, demonstrating that it is a topic of increasing interest. In 2020 there was a significant decrease in contributions on the subject, notwithstanding, compared to the results of the other keywords it is clear that it has a wide field of research.

On the other hand, referring to the issue of data mining/ desertion is the one that fewer publications has had over the years and although a slight increase in 2019 is denotes being very high and as well as the previous topics in The year 2020 decreased the amount of productions in this regard.

4. METHODOLOGY

This article seeks to show the literary review on the subject of data mining and university desertion, for this initially the keywords were selected, the review of publications referring to them was carried out and filtered in the online database IEEE XPLORE DIGITAL Library, taking into account the range of years from 2017 to 2020, finding the following results in summary:

Table 1: Key Words and Number of Results

Key Words	Search Results
School Dropout	214
Educational data mining	1.307
Data mining/desertion	79

For the development of literary analysis, 32 publications were selected for the topic of school dropout, 33 for educational data mining and 25 for data mining/desertion, taking into account that, for this last topic the results that provided the database used were not abundant and when making the filter of those publications that were directly related to the range of interest of this article, it was evidenced that the selected items were the most appropriate.

As the found publications for each topic were reviewed, a classification of these was performed and the most important information of each of them was extracted taking into account characteristics such as: title, year, summary, outcomes and conclusions.

5. RESULTS

Taking into account the information filter about the proposed keywords, it was evidenced that for the topic of student dropout (Figure 2) a total of 214 productions among which are 189 conferences, 3 magazines and 22 items were located; For each year from 2017 to 2020, an increase in the number of publications carried out on the subject is evident, however, for the last year a decrease is reflected in terms of conferences and magazines but in relation to articles increased slightly.

According to the graph of Figure 3, in relation to the issue of educational data mining, the productions were much more evident since a total of 1,307 was found on this content, the conferences were those that had a greater number of results in the search with a total of 1,147, followed by 157 items and in less quantity 3 published magazines during the 4 years in which the search was filtered.

Figure 2: Publications on School Dropout

Figure 3: Publications on Educational Data Mining.

For the found outcomes regarding data mining and dropout (Figure 4) the publications were very few in general since 97 were found in total, including 79 conferences and 18 items, there were no results for magazines.

Among the 90 chosen publications for this investigation (Figure 5), it was found that the year 2018 was in which more related publications were related, followed by 2017, 2016 and 2020. This last mentioned year represents the least number of publications related to the selected keywords for this literary review.

Figure 4: Publications on Data Mining/Desertion

Likewise, when analyzing the origin of these publications or their action field, it is found that only two of them are from Colombia and the rest (88) are studies and investigations of places in Asia, Brazil, United States, among others.

Figure 5: Number of Publications per Year.

Table 2: Describes the Authors that within the Selection of Chosen Publications for this Research made Several Contributions with their Productions

Author	Number of Publications	Description
Ahmed, S. A	3	- Ahmed, S. A., Billah, M. A., & Khan, S. I. (2020, July). - Ahmed, SA y Khan, SI (julio de 2019). - Ahmed, S. A., Billah, M. A., Al Noman, A., Doha, A. H., & Khan, S. I. (2020, July).
Medeiros, F.	2	- Brito, M., Medeiros, F., & Bezerra, E. (2019, July). - Santos, A. y de Medeiros, FPA (junio de 2020).
Pereira	2	- Pereira, RT y Zambrano, JC (2017, diciembre). - Pereira, MAG y Pedro, LFMG (junio de 2020).
Grandón, EE	2	- Pérez, A., Grandón, EE, Caniupán, M. y Vargas, G. (noviembre de 2018). - Medina, E. C., Chunga, C. B., Armas-Aguirre, J., & Grandón, E. E. (2020, June).
Shiratori, N.	2	- Shiratori, N. (2017, July). - Shiratori, N. (2018, July).

Santos, G. A. S.	2	- Santos, G. A. S., Belloze, K. T., Tarrataca, L., Haddad, D. B., Bordignon, A. L., & Brandao, D. N. (2020, July). - Brandao, I. V., da Costa, J. P. C., Santos, G. A., Praciano, B. J., Júnior, F. C., & Júnior, R. T. D. S. (2019, October).
Hegde, V.	2	- Hegde, V. y HS, SR (2017, diciembre). - Hegde, V. y Prageeth, PP (enero de 2018).
Da Silva	2	- da Silva, D. R., de Lima Martins, S., & Maciel, C. (2017, October). - da Silva, VRLB, de Albuquerque Silva, F. y Burégio, V. (junio de 2019).
Holanda, M & de Carvalho Victorino, M.	2	- Martins, L. C. B., Carvalho, R. N., Carvalho, R. S., Victorino, M. C., & Holanda, M. (2017, December). - da Fonseca Silveira, R., Holanda, M., de Carvalho Victorino, M. y Ladeira, M. (2019, diciembre).
Singh, SN	2	- Bansal, R., Mishra, A. y Singh, SN (enero de 2017). - Roy, S. y Singh, SN (2017, enero).

Table 2. Authors with several publications in the search filter

On the other hand, it was found that in each of the revised publications there were specific keywords for each topic, among them we can find:

Keywords for School Dropout

Computer, science, student, performance, neuronal network, abandonment, predict, automatic learning, detection of drops, risk of student delay, plug, Moodle, extraction patterns, students dropout, decision trees, data removal, success factors, mining of educational data, models of prediction, autonomy, motivation, transmedia narrative, collaborative work, graduate education, emotional intelligence, team learning, end, RNN, prevent desertion, good change of state, preliminary state; abandonment pattern; logistic regression model, reduction of dimensionality; Selection of features; Extraction of features; Retention; EDM; Educational trajectories, late desertion, engineering education, process mining, diversity, high-rate courses, e-learning, abandonment behavior; student model; Bayesian networks, prediction of time series, characteristics engineering, convolved neuronal networks, precision, education, classification, random forest, gradient increase, emotions, motivation to study, study success, macro data, learning machine, university desertion.

Keywords for Educational Data Mining

Educational data mining, online education, big data in education, online learning platforms, data processing, student performance, analysis, framework, programming language, e-learning, investment classroom, analytics learning, measurement of errors, precision, Naïve Bayes, J48, multilayer perceptron, autonomous learning platform, login on behavior, resources, data, pattern recognition, WEKA, prediction, education, action, class unbalanced; returning, classification, multiclass, learning analysis, students at risk, learning management systems, academic student, association rule mining, learning machine, School Dropout rate, cluster grouping, classifier, K-NN, machine support, educational industry, educational data, electroerotion tools, educational risk, control, administration, quality, educational program, EP, geoinformation systems, GIS, educational institution, student academic performance, mini educational data, domain modeling, analysis, analysis of feelings, level of aspect, attention, characteristic selection algorithm, IDEB, educational policies, KNN, university desertion.

Keywords for Data Mining / School Dropout

Educational Data Mining (EDM), clustering, data collection, drop, tools, Data Mining (DM), business intelligence; data analysis; e-learning, rule-based framework, neural network, machine learning, academic success, computer education, cyber-physical-social systems, dropout deep learning model, heterogeneous data, high-order clusters, lifelong learning, professional learning, learning analysis, MOOC, analysis of persistence and graduation of university students, association patterns,

classification, selection of characteristics, prediction of dropout, generation of functions, selection of functions, predictive analysis, decision tree, correspondence analysis, visualization data, distance higher education, drop out, active learning, group based sampling, entropy margin search, college dropout, deep learning, avoidance, school desertion, distance learning, accuracy, balanced accuracy, educational data extraction, dependencies of functions, initialization coefficient, rules, algorithm of grid search, support vector machine, early warning systems, attribute selection, STEM, computer science, feature selection, cross validation.

The consulted publications were classified as follows:

Key word: Students' dropout. Number of publications: 32	
TITLES	AUTHOR / YEAR
A Machine Learning Approach to Performance and Dropout prediction in Computer Science: Bangladesh Perspective	Ahmed, S. A., Billah, M. A., & Khan, S. I. (2020, July).
A Machine Learning Approach to Predict the Engineering Students at Risk of Dropout and Factors Behind: Bangladesh Perspective	Ahmed, SA y Khan, SI (2019, July).
A model for the identification of students at risk of dropout at a university of technology	Lottering, R., Hans, R. y Lall, M. (2020, August).
A report-type plugin to indicate dropout risk in the Virtual Learning Environment Moodle	Brito, M., Medeiros, F., & Bezerra, E. (2019, July).
Application of decision trees for detection of student dropout profiles	Pereira, RT y Zambrano, JC (2017, December).
Applying data analysis to identify early indicators for potential risk of dropout in Computer Science Students	Böttcher, A., Thurner, V. y Häfner, T. (2020, April).
Applying data mining techniques to predict student dropout: a case study	Perez, B., Castellanos, C., & Correal, D. (2018, May).
Collaborative transmedia learning environments for the promotion of students' autonomy and motivation a study developed with students at risk of dropout and school failure	Pereira, MAG y Pedro, LFMG (2020, June).
Comparative analysis of prediction techniques to determine student dropout: logistic regression vs decision trees	Pérez, A., Grandón, EE, Caniupán, M. y Vargas, G. (2018, November).
Comparison of Student Emotional Intelligence with completed and dropout students in a special graduate educational program	Imai, S., Shimamura, M., & Terasawa, K. (2019, July).
Data mining in educational systems for effective student mentoring	Mehra, M., Kalbande, DR, Mankar, S. y Mutsaddi, S. (2019, December).
Derivation of student patterns in a preliminary dropout state and identification of measures for reducing student dropouts	Shiratori, N. (2018, July).
Dropout Detection using non-academic data	Dharmawan, T., Ginardi, H. y Munif, A. (2018, August).
Dropout prediction system to reduce discontinue study rate of information technology students	Limsathitwong, K., Tiwatthanont, K. y Yatsungnoen, T. (2018, May).
EvolvedTree: analyzing student dropout in universities	Santos, G. A. S., Belloze, K. T., Tarrataca, L., Haddad, D. B., Bordignon, A. L., & Brandao, D. N. (2020, July).
From lab to production: learnt lessons and real life challenges of an early student-dropout prevention system	Ortigosa, A., Carro, RM, Bravo-Agapito, J., Lizcano, D., Alcolea, JJ y Blanco, O. (2019).
Higher education student dropout prediction and analysis through educational data mining	Hegde, V. y Prageeth, PP (2018, January).
Identifying factors contributing to university dropout with sparse logistic regression	Hori, G. (2018, July).
Influence of student diversity on educational trajectories in engineering high-failure rate courses that lead to late dropout	Salazar-Fernandez, J. P., Sepúlveda, M., & Munoz-Gama, J. (2019, April).
Mining educational data to predict academic dropouts: a case study in blended learning course	Sukhbaatar, O., Ogata, K., & Usagawa, T. (2018, October).
Modeling dropout behavior patterns using Bayesian networks in small-scale private university	Shiratori, N. (2017, July).

MOOC dropout prediction using FWTS-CNN model based on fused feature weighting and time series	Zheng, Y., Gao, Z., Wang, Y. y Fu, Q. (2020).
Perspectives to predict dropout in university students with machine learning	Solís, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M. (2018, July).
Predicting the probability of student dropout through EMIR using data from current and graduate students	Murakami, K., Takamatsu, K., Kozaki, Y., Kishida, A., Bannaka, K., Noda, I. & Nakata, Y. (2018, July).
Predicting university dropout through data analysis	Asha, P., Vandana, E., Bhavana, E., & Shankar, K. R. (2020, June).
Predictive model to reduce the dropout rate of university students in Perú: Bayesian networks vs. decision trees	Medina, E. C., Chunga, C. B., Armas-Aguirre, J., & Grandón, E. E. (2020, June).
Student dropout prediction: A KMUTT case study	Tenpipat, W. y Akkarajitsakul, K. (2020, September).
Students' disruption in business computer at the University of Phayao, Thailand	Nuankaew, P., Nuankaew, W., Phanniphong, K., & Bussaman, S. (2020, March).
Study motivation and academic emotions in engineering students a case study in German higher education	Fritzsche, ES, Schlingensiepen, J. y Kordts-Freudinger, R. (2018, April).
Supervised learning in the context of educational data mining to avoid university students' dropout	Kelly, J. D. O., Menezes, A. G., de Carvalho, A. B., & Montesco, C. A. (2019, July).
Understanding the student dropout in distance learning	de Oliveira, M. M., Barwaldt, R., Pias, M. R., & Espíndola, D. B. (2019, October).
Deep learning for dropout prediction in MOOCs	Sun, D., Mao, Y., Du, J., Xu, P., Zheng, Q., & Sun, H. (2019, October).
Key words: Educational data mining	
Number of publications: 33	
Application of big data in education, data mining and analytical learning: literature review	Sin, K. y Muthu, L. (2015).
A brief analysis of the key technologies and applications of educational data mining on online learning platform	Zhang, W. y Qin, S. (2018, March).
A framework to analyze performance of student's in programming language using educational data mining	Hegde, V. y HS, SR (2017, December).
A survey on educational data mining [2014-2019]	Hicham, A., Jeghal, A., Sabri, A. y Tairi, H. (2020, June).
A systematic mapping study of educational technologies based on educational data mining and learning analytics	Doko, E. y Bexheti, LA (2018, June).
Analysis of educational data mining using classification	Jalota, C. y Agrawal, R. (2019, February).
Characterizing educational data mining	da Silva, VRLB, de Albuquerque Silva, F. y Burégio, V. (2019, June).
Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA.	Mhetre, V., & Nagar, M. (2017, July).
Data level approach for imbalanced class handling on educational data mining: multiclass classification	Pristyanto, Y., Pratama, I., & Nugraha, A. F. (2018, March).
Data mining analysis on student's academic performance through exploration of student's background and social activities	Kiu, C. C. (2018, October).
Data mining approach to the identification of at-risk students	Ho, L. C., & Shim, K. J. (2018, December).
Early warning system for academic using data mining	Santoso, LW (2018, October).
Educational data analytics using association rule mining and classification	Rojanasvasu, P. (2019).
Educational data mining: a comprehensive study	Raju, R., Kalaiselvi, N., Divya, I. y Selvarani, A. (2020, July).
Educational data mining: analysis of drop out of engineering majors at the UNB - Brazil	da Fonseca Silveira, R., Holanda, M., de Carvalho Victorino, M. y Ladeira, M. (2019, December).
Educational data mining and learning analysis	Mishra, A., Bansal, R. y Singh, SN (2017, January).
Educational Data Mining (EDM) as a model for students' evaluation in learning environment	Hidayat, N., Wardoyo, R., & Azhari, S. N. (2018, October).
Educational data mining perspectives within university big data environment	Stefanova, K. y Kabakchieva, D. (2017, June).

Educational data mining to improve learner performance in Gauteng primary schools	Ramaphosa, K. I. M., Zuva, T., & Kwuimi, R. (2018, August).
Educational data mining tools: a survey from 2001 to 2016	Venkatachalapathy, K., Vijayalakshmi, V. y Ohmprakash, V. (2017, February).
Emerging trends in applications of big data in educational data mining and learning analytics	Roy, S. y Singh, SN (2017, January).
Literature review- techniques and algorithms used for various applications of Educational Data Mining (EDM).	Rambola, R. K., Inamke, M., & Harne, S. (2018, December).
Management of educational risk on the basis of data mining in SIG	Taganov, A. I., Kolesenkov, A. N., Perepelkin, D. A., & Zhuravlev, D. S. (2017, September).
Educational data mining to analyze students' behavior and performance	Kurdi, MM, Al-Khafagi, H. y Elzein, I. (2018, November).
Mining Educational Data to Predict Academic Dropouts: a Case Study in Blended Learning Course	Sukhbaatar, O., Ogata, K., & Usagawa, T. (2018, October).
Educational data mining for analyzing students' overall performance	Bansal, R., Mishra, A. y Singh, SN (2017, January).
Multi-attention fusion modeling for sentiment analysis of educational big data	Zhai, G., Yang, Y., Wang, H. y Du, S. (2020).
Performance analysis of feature selection algorithm for educational data mining	Zaffar, M., Hashmani, M. A., & Savita, K. S. (2017, November).
Relationship of federal funding to BID results in a state in Brazil: an approach based on educational data mining	Santos, A. y de Medeiros, FPA (2020, June).
Supervised learning in the context of educational data mining to avoid university students' dropout	Kelly, J. D. O., Menezes, A. G., de Carvalho, A. B., & Montesco, C. A. (2019, July).
The research and practice of a five-sided educational data mining framework	Zeng, T. (2017, October).
Application of educational data mining on analysis of students' online learning behavior	Jie, W., Hai-yan, L., Biao, C., & Yuan, Z. (2017, June).
Educational data mining: classifier comparison for the course selection process	Srivastava, S., Karigar, S., Khanna, R., & Agarwal, R. (2018, July).
Key word: Data Mining/ School dropout	
Number of publications: 25	
Estimating the school dropout trend by using data mining methods	Elbir, A., Gündüz, E. y Diri, B. (2018, October).
Classification of the province based on dropout rate using C4.5 Algorithm	Pertiwi, A. G., Widyaningtyas, T., & Pujiyanto, U. (2017, November).
A case study of applying the classification task for students' performance prediction	Guerra, M. S., Neto, H. A., & Oliveira, S. A. (2018).
A high-order clustering algorithm based on dropout deep learning for heterogeneous data in cyber-physical-social systems	Bu, F. (2017).
A learning analytics tool for predictive modeling of dropout and certificate acquisition on MOOCS for professional learning	Cobos, R., & Olmos, L. (2018, December).
A multifaceted data mining approach to understanding what factors lead college students to persist and graduate	Gopalakrishnan, A., Kased, R., Yang, H., Love, M. B., Graterol, C., & Shada, A. (2017, July).
An Integrated Framework with feature selection for dropout prediction in massive open online courses	Qiu, L., Liu, Y., & Liu, Y. (2018).
Classification and predictive analysis of educational data to improve the quality of distance learning courses	Brandao, I. V., da Costa, J. P. C., Santos, G. A., Praciano, B. J., Júnior, F. C., & Júnior, R. T. D. S. (2019, October).
Determination of dropout student profile based on correspondence analysis technique	Barros, T. M., Silva, I., & Guedes, L. A. (2019).
Different machine learning models to predict dropouts in MOOCs	Kashyap, A., & Nayak, A. (2018, September).
Early dropout prediction in distance higher education using active learning	Kostopoulos, G., Kotsiantis, S., Ragos, O., & Grapsa, T. N. (2017, August).
Early prediction of college attrition using data mining	Martins, L. C. B., Carvalho, R. N., Carvalho, R. S., Victorino, M. C., & Holanda, M. (2017, December).
Factors to predict dropout at the universities: A case of study in Ecuador	Mayra, A., & Mauricio, D. (2018, April).

Identification and systematization of indicatives and data mining techniques for detecting evasion in distance education	da Silva, D. R., de Lima Martins, S., & Maciel, C. (2017, October).
Identification of dropout students using educational data mining	Tasnim, N., Paul, MK y Sattar, AS (2019, February).
Improvement of weight backpropagation with particle swarm optimization for student dropout prediction	Sari, EY y Sunyoto, A. (2019, November).
Predicting and reducing dropout in virtual learning using machine learning techniques: a systematic review	Tamada, M. M., de Magalhães Netto, J. F., & de Lima, D. P. R. (2019, October).
Predicting the probability of student's degree completion by using different data mining techniques	Ahuja, R. y Kankane, Y. (2017, December).
Student pass rates prediction using optimized support vector machine and decision tree	Ma, X., & Zhou, Z. (2018, January).
Students' success predictive models based on selected input parameters set	Milinković, S., & Vujović, V. (2019, March).
Using ensemble decision tree model to predict student dropout in computing science	Venkatachalapathy, K., Vijayalakshmi, V. y Ohmprakash, V. (2017, February).
Data mining and its application in EDM domain	Chaturvedi, M. (2017, June).
A framework to establish a Rule-based specialized dropout prevention scheme	Ahmed, S. A., Billah, M. A., Al Noman, A., Doha, A. H., & Khan, S. I. (2020, July).
Towards the integration of business intelligence tools applied to educational data mining	Villegas-Ch, W., Luján-Mora, S. y Buenaño-Fernandez, D. (2018, March).
Undergraduate students' effectiveness in an institution with high dropout index	Oliveira, JL, Ambrósio, AP, Silva, U., Brancher, J. y Franco, JJ (2020, October).

Table 3. Articles consulted by keywords

6. CONCLUSIONS

Based on the current study and for the purpose of this article, it was possible to evidence that school dropout has been a topic of growing interest in different countries due to its great implications at a social and economic level, therefore, the need has arisen to find methods that allow reduce this type of issue and Educational Data Mining is considered as a step to achieve that objective. Nevertheless, the literature and research related to this topic are extensive separately, but when conducting the search combining the two topics, that is, connecting student dropout with Data Mining, it is clear that the results are more limited, this shows which is a field of research that has wide possibilities to continue exploring and implementing.

It was found that the main trend of the authors was related to the production of conferences, since in each search of the keywords in the online database IEEE Xplore Digital Library it was highlighted that this type of publication had a percentage of 88, 3% of the total results, followed by articles with 12.2% and academic journals with 0.3%.

After carrying out the literary review, 4 major themes were highlighted that were interconnected between the publications, one of them was prediction with 40 productions, abandonment with 44, data mining with 56 and desertion with 59. Likewise, the authors have addressed different data mining techniques for the prevention and prediction of student dropout, highlighting the great contribution of data mining in different fields of action, including education, which is entering and has wide possibilities.

As only 90 publications were selected in total, it is important to conduct a broader literary review, taking into account the publications that were not included in this research.

7. REFERENCES

1. Kelly, J. D. O., Menezes, A. G., de Carvalho, A. B., & Montesco, C. A. (2019, July). Supervised learning in the context of educational data mining to avoid university students dropout. In 2019 IEEE 19th International Conference on Advanced Learning Technologies [ICALT] (Vol. 2161, pp. 207-208). IEEE.

2. P.Rajendiran, K.R.Sekar, B.Sinivasan & R.Manikandan, "Customer Relationship Management in the Manufacturing Industry using Data Mining Techniques", *International Journal of Educational Science and Research (IJESR)*, Vol. 7, Issue 6, pp, 63-70
3. Asha, P., Vandana, E., Bhavana, E., & Shankar, K. R. (2020, June). Predicting University Dropout through Data Analysis. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)*(48184) (pp. 852-856). IEEE.
4. Snezhana Sulova, "Analysis and Evaluation on Online Shops Customers Through Data Mining Methods", *International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC)*, Vol. 8, Issue 1, pp, 35-44
5. Pereira, RT y Zambrano, JC (2017, December). Application of decision trees for detection of student dropout profiles. In *2017, 16th IEEE International Conference on Applications and Machine Learning (ICMLA)* (pp. 528-531). IEEE.
6. Gurpreet Singh & ER. Rajwinderkaur, "Review Paper on Decision Tree Data Mining Algorithms to Improve Accuracy in Identifying Classified Instances using Large Dataset", *International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)*, Vol. 7, Issue 4, pp, 67-70
7. Hori, G. (2018, July). Identifying Factors Contributing to University Dropout with Sparse Logistic Regression. In *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 430-433). IEEE.
8. Udayakumari Vidhyasagara Menon & Muhammed Refeque, "Big Data Analytics: Implications on Economic Planning and Implementation", *International Journal of Business and General Management (IJBGM)*, Vol. 5, Issue 5, Aupp; 59-64
9. Perez, B., Castellanos, C., & Correal, D. (2018, May). Applying data mining techniques to predict student dropout: a case study. In *2018 IEEE 1st colombian conference on applications in computational intelligence (colcaci)* (pp. 1-6). IEEE.
10. Vishal Jain, Gagandeep Singh Narula & Mayank Singh, "Implementation of Data Mining in Online Shopping System Using Tanagra Tool", *International Journal of Computer Science and Engineering (IJCSE)*, Vol. 2, Issue 1, pp, 47-58
11. Páramo, G., & Correa Maya, C. (1999). Deserción estudiantil universitaria. Conceptualización. *Revista Universidad EAFIT*, 35(114), 65-78.
12. S. V. S. Ganga Devi, "A Survey on Distributed Data Mining and its Trends ", *IMPACT: International Journal of Research in Engineering & Technology (IMPACT: IJRET)*, Vol. 2, Issue 3, pp, 107-120
13. Stefanova, K. y Kabakchieva, D. (2017, June). Educational data mining perspectives within university big data environment. In *2017 International Conference on Engineering, Technology and Innovation (ICE / ITMC)* (pp. 264-270). IEEE.
14. Zhang, W. y Qin, S. (2018, March). A brief analysis of the key technologies and applications of educational data mining on online learning platform. In *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)* (pp. 83-86). IEEE.
15. Rojanavas, P. (2019). Educational data analytics using association rule mining and classification. In *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)* (pp. 142-145). IEEE.
16. Sin, K. y Muthu, L. (2015). Application of big data in education, data mining and analytical learning: literature review. *ICTACT Magazine on Soft Computing*, 5 (4).
17. Pereira, RT y Zambrano, JC (2017, December). Application of decision trees for detection of student dropout profiles. In *2017, 16th IEEE International Conference on Applications and Machine Learning (ICMLA)* (pp. 528-531). IEEE.
18. Murakami, K., Takamatsu, K., Kozaki, Y., Kishida, A., Bannaka, K., Noda, I. & Nakata, Y. (2018, July). Predicting the Probability of Student Dropout through EMIR Using Data from Current and Graduate Students. In *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 478-481). IEEE.

19. Zheng, Y., Gao, Z., Wang, Y. y Fu, Q. (2020). *MOOC dropout prediction using FWTS-CNN model based on fused feature weighting and time series*. *Access IEEE*, 8, 225324-225335.
20. Tenpipat, W. y Akkarajitsakul, K. (2020, September). *Student dropout prediction: A KMUTT case study*. In *2020 1st International Conference on Big Data Analytics and Practices (IBDAP)* (pp. 1-5). IEEE.
21. Ahmed, S. A., Billah, M. A., & Khan, S. I. (2020, July). *A Machine Learning Approach to Performance and Dropout prediction in Computer Science: Bangladesh Perspective*. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
22. Ahmed, SA y Khan, SI (2019, July). *A Machine Learning Approach to Predict the Engineering Students at Risk of Dropout and Factors Behind: Bangladesh Perspective*. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
23. Lottering, R., Hans, R. y Lall, M. (2020, August). *A model for the identification of students at risk of dropout at a university of technology*. In *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1-8). IEEE.
24. Brito, M., Medeiros, F., & Bezerra, E. (2019, July). *A report-type plugin to indicate dropout risk in the virtual learning environment moodle*. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)* (Vol. 2161, pp. 127-128). IEEE.
25. Böttcher, A., Thurner, V. y Häfner, T. (2020, April). *Applying data analysis to identify early indicators for potential risk of dropout in Computer Science Students*. In *2020 IEEE Global Engineering Education Conference (EDUCON)* (pp. 827-836). IEEE.
26. Pereira, MAG y Pedro, LFMG (2020, June). *Collaborative transmedia learning environments for the promotion of students' autonomy and motivation a study developed with students at risk of dropout and school failure*. In *2020 XV Iberian Congress of Information Systems and Technologies (CISTI)* (pp. 1-6). IEEE.
27. Pérez, A., Grandón, EE, Caniupán, M. y Vargas, G. (2018, November). *Comparative analysis of prediction techniques to determine student dropout: logistic regression vs decision trees*. In *2018, 37th International Congress of the Chilean Society of Computer Sciences (SCCC)* (pp. 1-8). IEEE.
28. Imai, S., Shimamura, M., & Terasawa, K. (2019, July). *Comparison of Student EQ with completed and dropout students in a special graduate educational program*. In *2019, 8th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 414-419). IEEE.
29. Mehra, M., Kalbande, DR, Mankar, S. y Mutsaddi, S. (2019, December). *Data mining in educational systems for effective student mentoring*. In *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)* (pp. 1-5). IEEE.
30. Shiratori, N. (2018, July). *Derivation of student patterns in a preliminary dropout state and identification of measures for reducing student dropouts*. In *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 497-500). IEEE.
31. Dharmawan, T., Ginardi, H. y Munif, A. (2018, August). *Dropout Detection using non-academic data*. In *2018, IV International Conference on Science and Technology (ICST)* (pp. 1-4). IEEE.
32. Limsathitwong, K., Tiwatthanont, K. y Yatsungnoen, T. (2018, May). *Dropout prediction system to reduce discontinue study rate of information technology students*. In *2018, 5th International Business and Industrial Research Conference (ICBIR)* (pp. 110-114). IEEE.

33. Santos, G. A. S., Belloze, K. T., Tarrataca, L., Haddad, D. B., Bordignon, A. L., & Brandao, D. N. (2020, July). *EvolveDTree: Analyzing Student Dropout in Universities*. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 173-178). IEEE.
34. Ortigosa, A., Carro, RM, Bravo-Agapito, J., Lizcano, D., Alcolea, JJ y Blanco, O. (2019). *From lab to production: learnt lessons and real life challenges of an early student-dropout prevention system*. *IEEE Transactions on Learning Technologies*, 12 (2), 264-277.
35. Hegde, V. y Prageeth, PP (2018, January). *Higher education student dropout prediction and analysis through educational data mining*. In *2018, 2nd International Conference on Inventive Systems and Control (ICISC)* (pp. 694-699). IEEE.
36. Hori, G. (2018, July). *Identifying Factors Contributing to University Dropout with Sparse Logistic Regression*. In *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 430-433). IEEE.
37. Salazar-Fernandez, J. P., Sepúlveda, M., & Munoz-Gama, J. (2019, April). *Influence of student diversity on educational trajectories in engineering high-failure rate courses that lead to late dropout*. In *2019 IEEE Global Engineering Education Conference (EDUCON)* (pp. 607-616). IEEE.
38. Sukhbaatar, O., Ogata, K., & Usagawa, T. (2018, October). *Mining Educational Data to Predict Academic Dropouts: a Case Study in Blended Learning Course*. In *TENCON 2018-2018 IEEE region 10 conference* (pp. 2205-2208). IEEE.
39. Shiratori, N. (2017, July). *Modeling dropout behavior patterns using Bayesian networks in small-scale private university*. In *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 170-173). IEEE.
40. Solís, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M. (2018, July). *Perspectives to predict dropout in university students with machine learning*. In *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)* (pp. 1-6). IEEE.
41. Medina, E. C., Chunga, C. B., Armas-Aguirre, J., & Grandón, E. E. (2020, June). *Predictive model to reduce the dropout rate of university students in Perú: Bayesian Networks vs. Decision Trees*. In *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-7). IEEE.
42. Nuankaew, P., Nuankaew, W., Phanniphong, K., & Bussaman, S. (2020, March). *Students Disruption in Business Computer at the University of Phayao, Thailand*. In *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)* (pp. 95-100). IEEE.
43. Fritzsche, ES, Schlingensiepen, J. y Kordts-Freudinger, R. (2018, April). *Study motivation and academic emotions in engineering students a case study in German higher education*. In *2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 563-570). IEEE.
44. de Oliveira, M. M., Barwaldt, R., Pias, M. R., & Espíndola, D. B. (2019, October). *Understanding the Student Dropout in Distance Learning*. In *2019 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE.
45. Sun, D., Mao, Y., Du, J., Xu, P., Zheng, Q., & Sun, H. (2019, October). *Deep learning for dropout prediction in MOOCs*. In *2019 Eighth International Conference on Educational Innovation through Technology (EITT)* (pp. 87-90). IEEE.
46. Hegde, V. y HS, SR (2017, December). *A framework to analyze performance of student's in programming language using educational data mining*. In *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)* (pp. 1-4). IEEE.

47. Hicham, A., Jeghal, A., Sabri, A. y Tairi, H. (2020, June). A survey on educational data mining [2014-2019]. In 2020 International Conference on Intelligent Systems and Computer Vision (ISCV) (pp. 1-6). IEEE.
48. Doko, E. y Bexheti, LA (2018, June). A systematic mapping study of educational technologies based on educational data mining and learning analytics. In 2018 7th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-4). IEEE.
49. Jalota, C. y Agrawal, R. (2019, February). Analysis of educational data mining using classification. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 243-247). IEEE.
50. da Silva, VRLB, de Albuquerque Silva, F. y Burégio, V. (2019, June). Characterizing educational data mining. In 2019, XIV Iberian Congress of Information Systems and Technologies (CISTI) (pp. 1-5). IEEE.
51. Mhetre, V., & Nagar, M. (2017, July). Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA. In 2017 International Conference on Computing Methodologies and Communication (ICCMC) (pp. 475-479). IEEE.
52. Pristyanto, Y., Pratama, I., & Nugraha, A. F. (2018, March). Data level approach for imbalanced class handling on educational data mining multiclass classification. In 2018 International Conference on Information and Communications Technology (ICOIACT) (pp. 310-314). IEEE.
53. Kiu, C. C. (2018, October). Data Mining Analysis on Student's Academic Performance through Exploration of Student's Background and Social Activities. In 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA) (pp. 1-5). IEEE.
54. Ho, L. C., & Shim, K. J. (2018, December). Data mining approach to the identification of at-risk students. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 5333-5335). IEEE.
55. Santoso, LW (2018, October). Early warning system for academic using data mining. In 2018, Fourth International Conference on Advances in Computing, Communication and Automation (ICACCA) (pp. 1-4). IEEE.
56. Raju, R., Kalaiselvi, N., Divya, I. y Selvarani, A. (2020, July). Educational data mining: a comprehensive study. In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-5). IEEE.
57. da Fonseca Silveira, R., Holanda, M., de Carvalho Victorino, M. y Ladeira, M. (2019, December). Educational data mining: analysis of drop out of engineering majors at the UNB - Brazil. In 2019, 18th International IEEE Conference on Machine Learning and Applications (ICMLA) (pp. 259-262). IEEE.
58. Mishra, A., Bansal, R. y Singh, SN (2017, January). Educational data mining and learning analysis. In 2017, 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence (pp. 491-494). IEEE.
59. Hidayat, N., Wardoyo, R., & Azhari, S. N. (2018, October). Educational Data Mining (EDM) as a Model for Students' Evaluation in Learning Environment. In 2018 Third International Conference on Informatics and Computing (ICIC) (pp. 1-4). IEEE.
60. Ramaphosa, K. I. M., Zuva, T., & Kwuimi, R. (2018, August). Educational data mining to improve learner performance in Gauteng primary schools. In 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD) (pp. 1-6). IEEE.
61. Venkatachalapathy, K., Vijayalakshmi, V. y Ohmprakash, V. (2017, February). Educational data mining tools: a survey from 2001 to 2016. In 2017, Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM) (pp. 67-72). IEEE.

62. Roy, S. y Singh, SN (2017, January). *Emerging trends in applications of big data in educational data mining and learning analytics*. In 2017, 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence (pp. 193-198). IEEE.
63. Rambola, R. K., Inamke, M., & Harne, S. (2018, December). *Literature Review-Techniques and Algorithms Used for Various Applications of Educational Data Mining (EDM)*. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-4). IEEE.
64. Taganov, A. I., Kolesenkov, A. N., Perepelkin, D. A., & Zhuravlev, D. S. (2017, September). *Management of educational risk on the basis of data mining in GIS*. In 2017 International Conference" Quality Management, Transport and Information Security, Information Technologies"(IT&QM&IS) (pp. 577-580). IEEE.
65. Kurdi, MM, Al-Khafagi, H. y Elzein, I. (2018, November). *Educational data mining to analyze students' behavior and performance*. In 2018, the JCCO Joint International Conference on ICT in Education and Training, the International Conference on Computing in Arabic and the International Conference on Geocomputing (JCCO: TICET-ICCA-GECO) (pp. 1-5). IEEE.
66. Sukhbaatar, O., Ogata, K., & Usagawa, T. (2018, October). *Mining Educational Data to Predict Academic Dropouts: a Case Study in Blended Learning Course*. In TENCON 2018-2018 IEEE region 10 conference (pp. 2205-2208). IEEE.
67. Bansal, R., Mishra, A. y Singh, SN (2017, January). *Educational data mining for analyzing students' overall performance*. In 2017, 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence (pp. 495-497). IEEE.
68. Zhai, G., Yang, Y., Wang, H. y Du, S. (2020). *Multi-attention fusion modeling for sentiment analysis of educational big data*. Mining and Big Data Analysis, 3 (4), 311-319.
69. Zaffar, M., Hashmani, M. A., & Savita, K. S. (2017, November). *Performance analysis of feature selection algorithm for educational data mining*. In 2017 IEEE Conference on Big Data and Analytics (ICBDA) (pp. 7-12). IEEE.
70. Santos, A. y de Medeiros, FPA (2020, June). *Relationship of federal funding to BID results in a state in Brazil: an approach based on educational data mining*. In 2020, XV Iberian Congress of Information Systems and Technologies (CISTI) (pp. 1-4). IEEE.
71. Zeng, T. (2017, October). *The research and practice of a five-sided educational data mining framework*. In 2017, IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC) (pp. 1050-1053). IEEE.
72. Jie, W., Hai-yan, L., Biao, C., & Yuan, Z. (2017, June). *Application of educational data mining on analysis of students' online learning behavior*. In 2017 2nd International Conference on Image, Vision and Computing (ICIVC) (pp. 1011-1015). IEEE.
73. Srivastava, S., Karigar, S., Khanna, R., & Agarwal, R. (2018, July). *Educational Data Mining: Classifier Comparison for the Course Selection Process*. In 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE) (pp. 1-5). IEEE.
74. Elbir, A., Gündüz, E. y Diri, B. (2018, October). *Estimating the school dropout trend by using data mining methods*. In 2018, Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-4). IEEE.
75. Pertiwi, A. G., Widyaningtyas, T., & Pujianto, U. (2017, November). *Classification of province based on dropout rate using C4. 5 algorithm*. In 2017 International Conference on Sustainable Information Engineering and Technology (SIET) (pp. 410-413). IEEE.
76. Guerra, M. S., Neto, H. A., & Oliveira, S. A. (2018). *A Case Study of Applying the Classification Task for Students' Performance Prediction*. IEEE Latin America Transactions, 16(1), 172-177.
77. Bu, F. (2017). *A high-order clustering algorithm based on dropout deep learning for heterogeneous data in cyber-physical-social systems*. IEEE Access, 6, 11687-11693.

78. Cobos, R., & Olmos, L. (2018, December). A learning analytics tool for predictive modeling of dropout and certificate acquisition on MOOCs for professional learning. In *2018 IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 1533-1537). IEEE.
79. Gopalakrishnan, A., Kased, R., Yang, H., Love, M. B., Graterol, C., & Shada, A. (2017, July). A multifaceted data mining approach to understanding what factors lead college students to persist and graduate. In *2017 Computing Conference* (pp. 372-381). IEEE.
80. Qiu, L., Liu, Y., & Liu, Y. (2018). An integrated framework with feature selection for dropout prediction in massive open online courses. *IEEE Access*, 6, 71474-71484.
81. Brandao, I. V., da Costa, J. P. C., Santos, G. A., Praciano, B. J., Júnior, F. C., & Júnior, R. T. D. S. (2019, October). Classification and predictive analysis of educational data to improve the quality of distance learning courses. In *2019 Workshop on Communication Networks and Power Systems (WCNPS)* (pp. 1-6). IEEE.
82. Barros, T. M., Silva, I., & Guedes, L. A. (2019). Determination of Dropout Student Profile Based on Correspondence Analysis Technique. *IEEE Latin America Transactions*, 17(09), 1517-1523.
83. Kashyap, A., & Nayak, A. (2018, September). Different machine learning models to predict dropouts in MOOCs. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 80-85). IEEE.
84. Kostopoulos, G., Kotsiantis, S., Ragos, O., & Grapsa, T. N. (2017, August). Early dropout prediction in distance higher education using active learning. In *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1-6). IEEE.
85. Martins, L. C. B., Carvalho, R. N., Carvalho, R. S., Victorino, M. C., & Holanda, M. (2017, December). Early prediction of college attrition using data mining. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1075-1078). IEEE.
86. Mayra, A., & Mauricio, D. (2018, April). Factors to predict dropout at the universities: A case of study in Ecuador. In *2018, IEEE Global Engineering Education Conference (EDUCON)* (pp. 1238-1242). IEEE.
87. da Silva, D. R., de Lima Martins, S., & Maciel, C. (2017, October). Identification and systematization of indicatives and data mining techniques for detecting evasion in distance education. In *2017 Twelfth Latin American Conference on Learning Technologies (LACLO)* (pp. 1-8). IEEE.
88. Tasnim, N., Paul, MK y Sattar, AS (2019, February). Identification of dropout students using educational data mining. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1-5). IEEE.
89. Sari, EY y Sunyoto, A. (2019, November). Improvement of weight backpropagation with particle swarm optimization for student dropout prediction. In *2019, 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. IEEE.
90. Tamada, M. M., de Magalhães Netto, J. F., & de Lima, D. P. R. (2019, October). Predicting and Reducing Dropout in Virtual Learning using Machine Learning Techniques: A Systematic Review. In *2019 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE.
91. Ahuja, R. y Kankane, Y. (2017, diciembre). Predicting the probability of student's degree completion by using different data mining techniques. In *2017, Fourth International Conference on Image Information Processing (ICIIP)* (pp. 1-4). IEEE.
92. Ma, X., & Zhou, Z. (2018, January). Student pass rates prediction using optimized support vector machine and decision tree. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 209-215). IEEE.

93. Milinković, S., & Vujović, V. (2019, March). *Students' Success Predictive Models Based on Selected Input Parameters Set*. In *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-6). IEEE.
94. Oliveira, JL, Ambrósio, AP, Silva, U., Brancher, J. y Franco, JJ (2020, October). *Undergraduate students' effectiveness in an institution with high dropout index*. In *2020 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE.
95. Chaturvedi, M. (2017, June). *Data mining and its application in EDM domain*. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 829-834). IEEE.
96. Ahmed, S. A., Billah, M. A., Al Noman, A., Doha, A. H., & Khan, S. I. (2020, July). *A framework to establish a Rule-based specialized dropout prevention scheme*. In *2020, 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
97. Villegas-Ch, W., Luján-Mora, S. y Buenaño-Fernandez, D. (2018, marzo). *Towards the integration of business intelligence tools applied to educational data mining*. In *2018, IEEE World Engineering Education Conference (EDUNINE)* (pp. 1-5). IEEE.