Revista ARACÊ
ISSN: 2358-2472

USE OF MACHINE LEARNING ALGORITHMS FOR UNIVERSITY MANAGEMENT: A SYSTEMATIC REVIEW

https://doi.org/10.56238/arev6n4-071

Submitted on: 11/05/2024 **Publication date:** 12/05/2024

Aline Pacheco Primão[1], Alexandre Marino Costa[2], Diego Rossa[3] and Leonardo Flach[4].

ABSTRACT
This research aims to carry out a systematic review on the application of Machine Learning (ML) in university management. The study allows us to analyze previous research on the use of the most used ML algorithms and highlighted as the best, for a discussion and agenda. The searches were made in the international scientific databases Scopus and Web of Science, using the keywords Machine Learning, University, Higher Education. A total of 32 articles were selected for the sample. The results indicate that most surveys use more than one ML algorithm to perform predictions and that the Support Vector Machine (SVM) is the algorithm highlighted in most surveys as the best performer. Another conclusion identified is that most of the articles assess the risk of school dropout, student academic performance, predict student results, and analyze dropout in order to avoid dropout, dropout, or attrition from courses by the student.

Keywords: University Management, Machine Learning, Algorithms.

[1] Federal University of Santa Catarina, Brazil, aline.pacheco.pr@gmail.com
[2] Federal University of Santa Catarina, marinocad@gmail.com
[3] Brazil
diegorossa67@gmail.com
[4] Federal University of Santa Catarina, Brazil, leonardo.flach@gmail.com

Revista ARACÊ
ISSN: 2358-2472

INTRODUCTION

Between 2009 and 2019, Brazil recorded a 43.7% increase in enrollment in higher education, and when we talk about federal institutions, this growth was 59.1% (INEP, 2019). Thus, the management of these organizations is of great relevance to achieve positive results in decision-making, in the search for correct public spending and university excellence.

University Management (GU) can be understood as the act of managing the university. According to Schlickmann (2013), university management is under construction, as it is still associated with administration and other fields such as education, with this, there is a diffusion of knowledge and theoretical bodies (Schlickmann, 2013). University Management is considered complex, as these institutions have predominant bureaucratic processes and undergo frequent changes due to political influences and interest groups (Gomes et al., 2013).

To collaborate with GU, Educational Data Mining (EDM) emerges, which, despite being an emerging field, has been gaining attention in recent years, as it can be used to generate information that helps in decision-making in the educational process (Sultana et al., 2017). However, educational institutions have a very large volume of academic data, and to produce the searches and analysis of this data, an investigation based on the extraction of knowledge is necessary, for this, Machine Learning (ML) algorithms are feasible options (Silva et al., 2020).

Machine Learning studies use statistical models, and have been used to predict the risk of an event occurring (Silva et al., 2020). Since, from the literature review, it is observed that an ML algorithm cannot obtain a good performance in all types of applications, and that it is often necessary to bring together more than one algorithm for performance gain, it is necessary to analyze the most used algorithms and verify the ones that best fit the research. Delen (2010) and Adekitan and Salau (2019) state that a set of algorithms perform better than individual models (Delen, 2010); (Adekitan & Salau, 2019).

From the previous introduction, this bibliometric research is justified: 1) To evaluate the production of Machine Learning research and what it brings to the University Management; 2) Evaluate the areas where EM techniques are most used; 3) Analyze the ML algorithms for related searches and those that bring better results for prediction within the GU.

Revista ARACÊ
ISSN: 2358-2472

The article is presented as follows: 1) Introduction, where a brief presentation of the researched theme and justification of this research was provided; 2) Methods, where the objective of the research is presented, how the review was carried out, search bases and criteria for selection and exclusion of articles; 3) Results, where the characteristics, synthesis and graph of the results of the articles are exposed; and 4) Discussion, where the relevant results, limitations, general interpretations and future implications are highlighted.

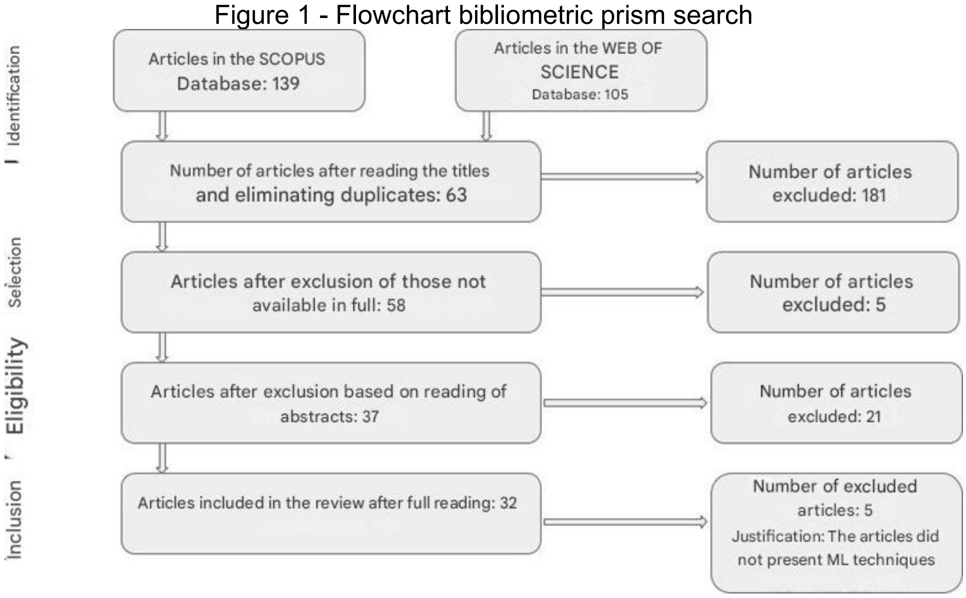
RESEARCH METHOD

In order to analyze previous research on the proposed theme, or related themes and to make a discussion of the literature, a bibliometric search was carried out from the following keywords: Machine Learning AND University AND Higher Education in the Scopus and Web of Science databases from the year 2010. The search was conducted in December 2020, in all languages.

244 articles were found, removing the repeated ones and those that did not fit the search, after reading the titles, 63 articles remained. Soon after, those that were not available in full were removed, leaving 58 articles. Following the reading of the abstracts, those that were not presented as a focus on university management were excluded, leaving 37 articles. And finally, after reading in full, 5 were excluded because they did not present the Machine Learning algorithms, thus leaving a total of 32 articles.

This research used the Prism recommendation of Moher et al. (2009) and presented the flowchart in Figure 1.

Figure 1 - Flowchart bibliometric prism search



Source: Prepared by the authors.

Revista ARACÊ
ISSN: 2358-2472

In section 3, the results of the bibliometric research will be addressed, bringing the characteristics of the selected articles, syntheses and graphs.

DISCUSSION AND ANALYSIS OF RESULTS

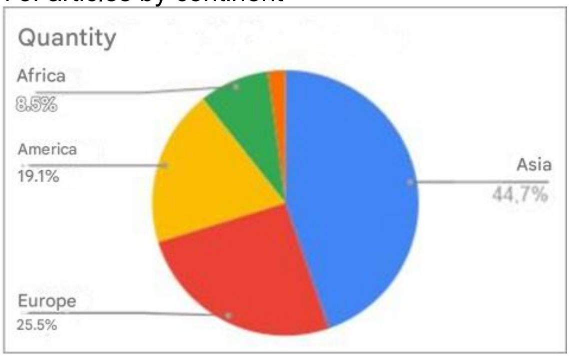
With the research, it was observed that the years 2019 and 2020 were the most evident, providing 26 of the 32 articles. It was also found that 29 of these were in English, 2 in Spanish and 1 in Portuguese.

The Asian continent is the most significant in publications with 21 representations, with China being portrayed in 5 articles and Taiwan in 4, followed by Europe with 12 and America with 9 representations. It is observed that an article can be represented by more than one country or continent.

Although only one of the studies is in Portuguese, Brazil is represented in 3 articles. Costa et al. (2017) stands out, in which on the date of the search in the databases, there were 113 citations. Figure 2 represents the arrangement of the articles by continent.

Figure 2 - Breakdown of articles by continent

Continent	Amount
Asia	21
Europe	12
America	9
Africa	4
Oceania	1

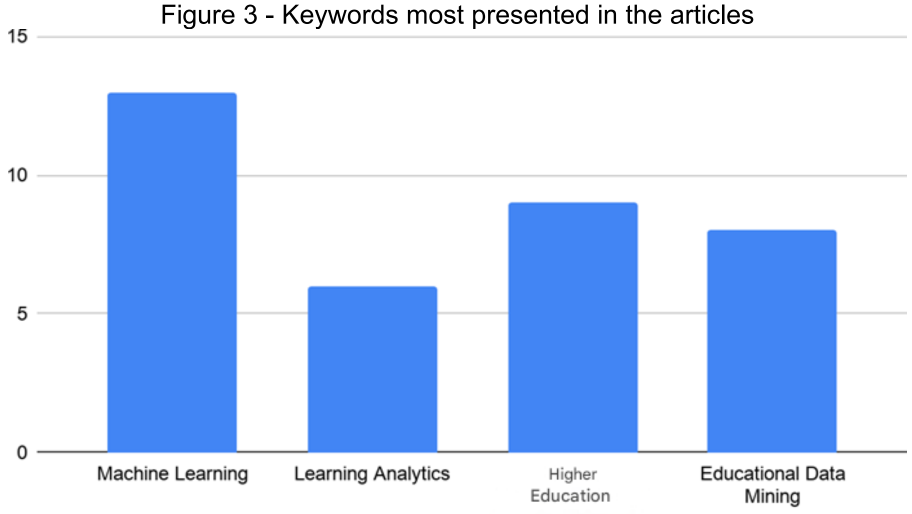


Source: Prepared by the authors (2021).

The most used keywords were Machine Learning with 13 citations, Higher Education/Educação Superior/Educación Superior with 9 citations, Educational Data Mining with 8 citations, and Learning Analytics with 6 citations. Figure 3 represents the number of times the keyword was cited in an article.

Revista ARACÊ
ISSN: 2358-2472

Figure 3 - Keywords most presented in the articles



Source: Prepared by the authors (2021).

Regarding the objectives of the articles, 6 of them have as their main objective to predict students' academic results, 6 to analyze dropout to avoid dropout, dropout or attrition, 5 to predict students at risk of failure, 2 to predict student grades, 2 to analyze student procrastination, 2 to predict academic performance, 1 to review texts in student opinion surveys, 1 predict the number of students with low engagement in courses, 1 propose a system to predict path for students in the college preparatory year, 1 explore the learning behaviors generated by students, 1 propose a predictive model with improved accuracy rates from the analysis of student data, 1 study what types of prediction models perform best, and 1 promote the intuitive use of dropout analysis techniques. Table 1 shows the articles by similar objectives.

Table 1 - Articles by similar objectives

OBJECTIVES	QUANTITY
Analyze student procrastination	2
Predict student grades	2
Predict students at risk of failure	5
Predict student academic outcomes	6
Proofreading texts in student surveys	1
Evaluate technical data from students	1
Know the actual number of students on a platform	1
Analyze dropout to avoid dropout, dropout, or attrition	6
Predict the number of students with low engagement in courses	1
Propose a system to predict a path for students in the preparatory year of college	1
Explore student-generated learning behaviors	1
Propose a predictive model with improved accuracy rates from student data analysis	1
Predict academic performance	2
Study which types of forecasting models perform best	1

Revista ARACÊ
ISSN: 2358-2472

Promote the intuitive use of evasion analysis techniques	1

Source: Prepared by the authors.

With the related research, we found that one of the major concerns of University Management is school dropout, especially when dealing with public institutions, as they need to guarantee relevant results and affirm the graduation of students for the job market (Andifes, 1996). Likewise, being able to predict school performance, allowing institutions to develop early actions to help students improve their grades and, consequently, train more qualified professionals and reduce school dropout is very important for higher education institutions.

For Delen (2010), high school dropout rates affect enrollment planning and bring an overload of work to recruit new students, while for students, dropping out before obtaining a diploma represents that human potential has not been explored and also causing a low return on the institution's investments (Delen, 2010). For Lau (2003 apud Delen, 2010) one of the probable causes of school dropout may be the difficulty of adapting to the university, affecting the institution's classification, reputation and financial (Lau, 2003 apud Delen, 2010).

Regarding the temporality of the data used for the research, a great distinction is perceived. Delen (2010), for example, used 5-year data from freshmen at a university to manage student retention, while Sultana et al. (2017) analyzed first-year college students in electrical engineering and computer science to predict performance. Ezz and Elshenawy (2019) analyzed the behavior of students in the college preparatory course, Hai-Tao et al. (2020) developed a model to predict student performance even before the course started, Deo et al. (2020) used data from 6 years in multiple courses to investigate and propose a performance prediction model, Adekitan and Salau (2019) used data to analyze performance in the first 3 years of graduation to be able to predict the final result of students, while Costa et al. (2017) used 2 types of data from students who took 10 and 16-week introductory programming courses to predict early student failure.

MACHINE LEARNING ALGORITHMS

Of the most used algorithms in the searches, 18 articles used Support Vector Machine (SVM), 16 Random Forest (RF), 16 Decision Tree (DT), 15 Neural Network (NN), 10 Logistic Regression (LR), 6 K-Nearest Neighbor (KNN) and 5 Multi-Layer Perceptron

Revista ARACÊ
ISSN: 2358-2472

(MLP). Table 2 shows the number of times Machine Learning algorithms were used in surveys, where most surveys used more than one algorithm.

Table 2 - Machine learning algorithms in the articles

ALGORITHM	QUANTITY
SVM - Support Vector Machine	18
RF - Random Forest	16
DT - Decision Tree	16
LR - Logistic Regression	10
ANN - Artificial Neural Network or NN - Neural Network or DNN - Deep Neural Network or PNN - Probabilistic Neural Network	15
KNN - K-Nearest Neighbor	6
MLP - Multilayer Perceptron	5

Source: Prepared by the authors.

Delen (2010) used Artificial Neural Network, Decision Tree, Support Vector Machine and Logistic Regression to perform the prediction, and SVM was the algorithm that had the best performance. The author states that balanced data brought better performances than unbalanced data, regardless of the algorithm used. Another observation by Delen (2010) is that Decision Tree algorithms provide a more transparent view of where and how they do it compared to Support Vector Machine (Delen, 2010).

Adekitan and Salau (2019) applied six algorithms (Probabilistic Neural Network, Random Forest, Decision Tree, Naive Bayes, Tree Ensemble and Logistic Regression) independently to predict the Cumulative Average of Final Grades of students with data from the first three years of graduation, the approach that presented the best performance was Logistic Regression (Adekitan & Salau, 2019). Finally, the authors combined all the algorithms into one model to get the benefits of each together. The authors used high school grades, class participation level, attendance, intermediate grades, lab reports, homework assignment grades, seminar scores, assignment completion, and overall grades to predict truancy in higher education (Adekitan & Salau, 2019).

Silva et al. (2020) empregou Regressão Logística, K-Nearest Neighbors, Naive Bayes, Support Vector Machines, Decision Tree Based Methods (C trees, Baggins, Random Forest, Boosting) e Penalized Methods (Ridge, Lasso, Elastic Net) para avaliar as melhores variáveis para a performance de modelos.

Costa et al. (2017) selected the Naive Bayes, Decision Tree, Artificial Neural Network and Support Vector Machine algorithms for their study, because according to the author these techniques have good efficacy in different environments and have been used to identify students with academic failures (Costa et al., 2017). The results obtained with the algorithms were similar, but the Support Vector Machine was the one that achieved the best effectiveness in the research (Costa et al., 2017).

Gamie et. al. (2020) used ML algorithms together to analyze the factors that impact students' performance in the last year of the course and propose a predictive model with improved accuracy rates compared to others in the same dataset. The authors also evaluated the most significant characteristics that can affect student performance. The model by Gamie et al. (2020) followed the following steps: initialize resource groups based on data sources; generate possible combinations of repeating groups in order to detect the best combination, Apply Support Vector Machine, Decision Tree and Neural Networks to each partition (combination of features); Apply Random Forest as a bagging technique in each partition (combination of resources); Apply XgBoost and AdaBoost with Decision Tree as the base learner of each partition (combination of resources); select the best classifier along with the best combination of groups; Increase the Support Vector Machine linear and non-linear and the Random Forest and save the results and; compare the accuracy of the classification (Gamie et al., 2020). In this research, it was concluded that demographic data do not affect the accuracy of the final results, so they were excluded so as not to generate unnecessary computational effort (Gamie et. al., 2020).

Few authors of the selected research used only one algorithm for the study, as is the case of Beaulac and Rosenthal (2019), who used Random Forest to predict whether the student will complete their undergraduate degree and to predict which courses attract students the most. The authors reiterate that RF is easy to use, quick to train, and outperforms Linear Regression models in prediction accuracy. Beaulac and Rosenthal (2019) adjusted the RF classifiers and compared them with two logistic regression models to predict whether a student completes their course or not (Beaulac & Rosenthal, 2019).

Another author who took advantage of only one algorithm was Chui et al. (2020), who proposed a reduced evaluation model with the Support Vector Machine to predict at-risk students and potential students dropping out of their courses. Afterwards, the author performed a comparison with related research, thus, he concluded that his model can be adopted, as it reduces the training time when the dataset provided is large (Chui et al., 2020).

Nikolic et al. (2020) presented a system based on Natural Language Processing (NLP) to evaluate free texts expressed by students in opinion polls in the Serbian language and make a higher institution able to verify which aspects students are satisfied or

Revista ARACÊ
ISSN: 2358-2472

dissatisfied with. The author also used the K-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machine algorithms to perform machine learning and the one that obtained the best evaluation was the Support Vector Machine (Nikolic et al., 2020).

Support Vector Machine, Naive Bayes, Logistic Regression, JRip, J48, Multilayer Perceptron, and Random Forest was utilized by Mia et al. (2019) to predict the number of students registered in a semester to help with the pre-planning of a private Bangladeshi university. Mia et al. (2019) compared the classifiers and the Support Vector Machine obtained better accuracy, while Random Forest obtained lower accuracy. The author intends to consider more attributes and increase the number of universities evaluated in the future (Mia et al., 2019).

Suharjito (2019) explored the K-Nearest Neighbor, Naive Bayes, and Decision Tree algorithms to analyze and find a better modeling solution in identifying predictors of student dropout at a Jakarta university. By combining the algorithms with Ensemble Classifier method and tested several times the accuracy performed better. Among the algorithms, the one that obtained the best result was the K-Nearest Neighbor (Suharjito, 2019).

DATA COLLECTION

Regarding data collection, most of the studies used only the collection in the databases of the academic systems with representation of 26 articles, the studies that used data collection from the academic systems together with one or more questionnaires add up to 4 articles, and the studies that took into account only questionnaires add up to 2 articles.

Vidhya and Vadivu (2020) used a set of questionnaires covering all aspects of students' learning factors (students' personal data, learning pattern, behavior, emotional factors, multiple intelligence, and cognitive abilities). From the results obtained in the questionnaires, the authors applied the machine learning algorithms and thus managed to classify the students in the categories "Excellent", "Good", "Average" and "Poor". With the results, it is possible to take measures to improve the students' result and also the reputation of the institution (Vidhya & Vadivu, 2020).

Muñiz et al. (2019) used in a first step the data extracted from the institutional system where he extracted: identification data, gender, place of birth, nationality, disability, family size, parents' qualifications and current occupations, average high school grade, university entrance exam score, age at admission, date of first enrollment, priorities indicated in the course admission application, area of knowledge corresponding to the

Revista ARACÊ
ISSN: 2358-2472

student's course, number of credits enrolled, past credits and average score, scholarship, current academic situation and transfer destination, if any. In a second stage, a questionnaire was carried out related to marital status, income level, type of housing during the course, motivation for choosing the course and university, participation in welcome activities for freshmen and their opinion, time spent studying, working and housework, evaluation of the requirements of the satisfaction program with scores, evaluation of personal relationships, intention to drop out and reasons, satisfaction with the university and, if the student dropped out, the current situation and satisfaction with the results of his decision (Miñiz et al., 2019).

Delen (2010) used variables related to the academic level, financial and demographic characteristics of the students (Delen, 2010). Adekitan and Salau (2019) used high school grades, class participation level, attendance, intermediate grades, lab reports, homework assignment grades, seminar scores, assignment completion, and overall grades to predict school dropout in higher education (Adekitan & Salau, 2019).

Silva et al. (2020) used data made available by the Lattes Platform of the National Council for Scientific and Technological Development (CNPq) to analyze the data of the professors such as time of graduation, publications in the year, doctorate and exclusive dedication, along with institutional data, where the student's data such as grade in the entrance exam and mathematics grade in the entrance exam were taken. Silva et al. (2020) states that calculus subjects are related to the deficiency of general knowledge coming from basic education, so the use of mathematics grades can directly influence student performance in higher education (Silva et al., 2020).

According to Sultana et al. (2017), when compared between some ML algorithms (Logistic Regression, Decision Tree, Naive Bayes and Neural Networks), Neural Networks performed better in predicting student performance, but using only cognitive characteristics, so the authors evaluated cognitive and non-cognitive characteristics to predict school dropout. For Sultana et al. (2017), non-cognitive characteristics can help increase the accuracy of the prediction of dropout. The authors collected data through self-assessment questionnaires, social support questionnaire, skills and leadership questionnaire, and student self-assessment questionnaire to analyze the non-cognitive characteristics of first-year students in the faculty of electrical engineering and computer science. The data collected from the questionnaires were pre-processed together with secondary sources and Machine Learning algorithms (Decision Tree, Logistic Regression, Naive Bayes and Neural

Networks) were used to classify the data. Sultana et al. (2017) found that some non-cognitive characteristics help in prediction and accuracy and some other characteristics do not, depending on the course analyzed, and that when non-cognitive characteristics are coupled with cognitive characteristics they result in better predictions and accuracy. The objective of the research was to use ML algorithms to generate data and be able to inform students with poor performance in time to improve their performance and thus reduce school dropout (Sultana et al., 2017).

Gamie et al. (2019) analyzed a set of dimensions and each dimension included a set of variables. Forecast analysis is applied for each dimension, and finally, a comparison between the dimensions is performed. Three factors were the basis of the variables: student activities, teaching style, and content categorization.

Suharjito (2019) used demographic and academic performance data (cumulative grade point average, internal assessment, external assessment, extracurricular activities, high school transcript, and social activities to predict the student at risk of dropping out. According to Suharjito (2019), in the first two years, gender also influences the quality of learning, as well as characteristics such as age, financial constraints, student absence, parental influence, job opportunity, and marital status (Suharjito, 2019).

FINAL CONSIDERATIONS

The article aimed to make a systematic review in order to evaluate previous research that used Machine Learning techniques in university management.

The research found that within University Management, most of the articles represent an involvement with the risk of school dropout, student academic performance, being able to predict student results and analyze dropout in order to avoid dropout, abandonment or attrition of courses by the student. These results show that the concern of higher education institutions with school dropout is great on all continents.

It was also observed that the algorithms most used in the research were Support Vector Machine, Decision Tree, Random Forest, Neural Networks and Logistic Regression. The algorithm that stands out is Support Vector Machine, presenting better accuracy in most of the surveys. Decision Tree was considered the most transparent. Another one that brought good results was Random Forest, which can be easy to use and train. Although Naive Bayes is not among the most used algorithms in research, it obtained the best results in the work of Gamie et al. (2019).

Revista ARACÊ
ISSN: 2358-2472

Random Forest, Support Vector Machine, Naive Bayes, Decision Tree, and XGBoost were used individually in some research. Although some authors prefer to apply only one algorithm, in general, studies show that the use of a set of selected algorithms brings a better prediction of the results.

Brazil is represented in the research with 3 articles, highlighting the article by Costa et al. (2017) which has a large number of citations (113). However, the most accepted articles are limited to the English language.

Finally, this article helped to analyze the most used algorithms and those that perform best in Machine Learning research, the objectives of the research and to verify, in addition to other countries, what Brazil has developed on the subject.

It is intended to deepen the study of the ML algorithms most used in the evaluated research and analyze how they perform in school dropout in Brazilian higher education, analyze the factors that most impact school dropout in Brazilian federal institutions, and from this, use the algorithms that show effectiveness in related research for training and testing of the database of a federal institution in Brazil and propose a dropout prediction model for the Brazilian IFES.

Revista ARACÊ
ISSN: 2358-2472

REFERENCES

1. Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. Journal of Applied Research in Higher Education.

2. 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.

3. Andifes. (1996). Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas. Resumo do relatório apresentado à Andifes, Abruem e SESu/MEC pela Comissão Especial, 55-65.

4. Beaulac, C., & Rosenthal, J. S. (2019). Predicting university students' academic success and major using random forests. Research in Higher Education.

5. Chui, K., Fung, D., Lytras, M., & Lam, T. (2020). Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. Computer in Human Behavior.

6. 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 student's academic failure in introductory programming courses. Computers in Human Behavior, 247-256.

7. Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems.

8. Deo, R. C., Yaseen, Z. M., Al-Ansari, N., Nguyen-Huy, T., Langlands, T. A. M., Galligan, L. (2020). Modern artificial intelligence model development for undergraduate student performance prediction: An investigation on engineering mathematics courses. IEEE Access.

9. Ezz, M., & Elshenawy, A. (2019). Adaptive recommendation system using machine learning algorithms for predicting student's best academic program. Education and Information Technologies.

10. Fernández-Martín, T., Solís-Salazar, M., María Teresa, M. T., & Moreira-Mora, T. E. (2018). Un análisis multinomial y predictivo de los factores asociados a la deserción universitaria. Revista Eletrónica Educare/Educare Electronic Journal.

11. Fernández, D. B., Gil, D., & Mora, S. L. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. Sustainability.

Revista
ARACÊ
ISSN: 2358-2472

12. Freitas, F. A. da S., Vasconcelos, F. F. X., Peixoto, S. A., Hassan, M. M., Dewan, M. A. A., Albuquerque, V. H. C. de, & Rebouças Filho, P. P. (2020). IoT systems for school dropout prediction using machine learning techniques based on socioeconomic data. Electronics.

13. Gamie, E., El-Soud, M. S. A., & Salama, M. A. (2020). Comparative analysis for boosting classifier in the context of higher education. International Journal of Emerging Technologies in Learning.

14. Gamie, E. A., El-Soud, M. S., Salama, M. A., & Hussein, W. (2019). Multi-dimensional analysis to predict student's grades in higher education. International Journal of Emerging Technologies in Learning, 14(2).

15. Gomes, O. F., Gomide, T. R., Gomes, M. A. N., Araújo, D. C., Martins, S., & Faroni, W. (2013). Sentidos e implicações da gestão universitária para os gestores universitários. Revista Gestão Universitária na América Latina - GUAL.

16. Hai-Tao, P., Ming-Qu, F., Hong-Bin, Z., Bi-Zhen, Y., Jin-Jiao, L., Chun-Fang, L., Yan-Ze, Z., & Rui, S. (2020). Predicting academic performance of students in Chinese-foreign cooperation in running schools with graph convolutional network. Neural Computing and Applications.

17. Hung, H. C., Liu, I.-F., Liang, C.-T., & Su, Y.-S. (2020). Applying educational data mining to explore student's learning patterns in the flipped learning approach for coding education. Symmetry.

18. Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. Hindawi - Computational Intelligence and Neuroscience.

19. Iatrellis, O., Savvas, I. K., Fitsilis, P., & Gerogiannis, V. C. (2020). A two-phase machine learning approach for predicting student outcomes. Education and Information Technologies.

20. Inep. (2021). Censo da educação superior 2019. Available at: https://download.inep.gov.br/educacao_superior/censo_superior/documentos/2020/Press_Kit_Censo_Superior_2019.pdf. Last accessed January 20, 2021.

21. Inep. (2017). Metodologia de indicadores e trajetória de curso. Available at: https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2017/metodologia_indicadores_trajetoria_curso.pdf. Last accessed January 20, 2021.

22. Lau, L. K. (2003). Institutional factors affecting students retention. Education, 126-137.

23. Manzanares, M. C. S., Sánchez, R. M., González, A. A., Llamazares, M. C. E., & Deus, M. A. Q. (2018). Detección del alumno en riesgo en titulaciones de ciencias de la salud: Aplicación de técnicas de learning analytics. European Journal of Investigation in Health, Psychology and Education.

24. Mia, M., Biswas, A., Satter, A., & Habib, T. (2019). Registration status prediction of students using machine learning in the context of private university of Bangladesh. International Journal of Innovative Technology and Exploring Engineering.

25. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Available at: www.prisma-statement.org.

26. Muñiz, L. R., Bernardo, A. B., Esteban, M., & Diaz, I. (2019). Dropout and transfer paths: What are the risk profiles when analyzing university persistence with machine learning techniques? PLOS One.

27. Musso, M., Hernández, C., & Cascallar, E. (2020). Predicting key educational outcomes in academic trajectories: A machine-learning approach. Higher Education.

28. Nikolic, N., Grljevic, O., & Kovacevic, A. (2020). Aspect-based sentiment analysis of reviews in the domain of higher education. Electronic Library, 44-64.

29. Pillai, B. R. (2019). Deep regressor: Cross subject academic performance prediction system for university level students. International Journal of Innovative Technology and Exploring Engineering.

30. Schlickmann, R. (2013). Administração universitária: Desvendando o campo científico no Brasil. Universidade Federal de Santa Catarina. (Doctoral dissertation).

31. Shiratori, N. (2017). Modeling dropout behavior patterns using Bayesian networks in small-scale private university. In Proceedings of the 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Hamamatsu, Japan, 170-173.

32. Silva Filho, R. L. L. e, Motejunas, P. R., Hipólito, O., & Lobo, M. B. de C. M. (2007). A evasão no ensino superior brasileiro. Cadernos de Pesquisa, 37(132), 641-659.

33. Silva, A. F. da, Almeida, A. T. C. de, & Ramalho, H. M. de B. (2020). Predição do risco de reprovação no ensino superior usando algoritmos de machine learning. Teoria e Prática em Administração, July-December.

34. Suharjito, N. H. (2019). Predictive modelling of student dropout using ensemble classifier method in higher education. Advances in Science, Technology and Systems Journal, 4(4), 206-211.

35. Sultana, S., Khan, S., & Abbas, M. A. (2017). Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. International Journal of Electrical Engineering Education, 54(2), 105-118.

36. Tsai, S.-C., Chen, C.-H., Shiao, Y.-T., Ciou, J.-S., & Wu, T.-N. (2020). Precision education with statistical learning and deep learning: A case study in Taiwan. International Journal of Educational Technology in Higher Education.

Revista ARACÊ
ISSN: 2358-2472

37. Vidhya, R., & Vadivu, G. (2020). Towards developing an ensemble-based two-level student classification model (ESCM) using advanced learning patterns and analytics. Journal of Ambient Intelligence and Humanized Computing.

38. Wotaifi, T., & Al-Shamery, E. (2019). Mining of completion rate of higher education based on fuzzy feature selection model and machine learning techniques. International Journal of Recent Technology and Engineering.

39. Xiao, M., & Yi, H. (2020). Research on adaptive learning prediction based on XAPI. International Journal of Information and Education Technology.

40. Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y.-M., & Lim, H. (2020). Predicting course achievement of university students based on their procrastination behaviour on Moodle. Soft Computing.

41. Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y.-M., & Lim, H. (2020). Prediction of student's procrastination behaviour through their submission behavioural pattern in online learning. Journal of Ambient Intelligence and Humanized Computing.