Machine Learning Techniques For Statistical And Tensor Signal Processing Applied To Education: A Systematic Literature Review.

Francisco Antonio Nascimento[1], Francisco José da Costa[2], Samuel Soares de Souza Santos [3], Tiago Fernando Hansel[4], David Pereira de Araújo[5], Aldecio Machado dos Santos[6].

[1]*(Universidade Federal do Ceará (UFC) – Brasil)*
[2]*(Universidade Federal do Ceará (UFC) – Brasil)*
[3]*(Instituto Federal de Educação, Ciência e Tecnologia do Amazonas – IFAM, Brasil)*
[4]*(Universidade Federal do Paraná – UFPR- Brasl)*
[5]*(Universidade Federal de Uberlândia - UFU - , Brasil)*
[6]*(Universidade Alto Verde do Rio do Peixe – UNIARP-Brasil)*

Abstract: *Signal processing is the field of science that studies methods of analysis and synthesis of signals, sequences or measurements that can represent physical quantities. To do this, mathematical tools are used to describe the transformations of data and signals. The digital environment is growing every day, and because of this more and more research is being done on artificial intelligence and machine learning in several areas. One of the areas that needs more research and attention is education, which is an area that has a large volume of data and information. In the literature many authors are using these techniques to predict student performance, school dropouts, and also to create optimizations between teacher/student dynamics.*

Materials and Methods: *To understand the context presented, a systematic review of the literature was conducted in order to understand how the Signal Processing and Machine Learning is applied in the area of Education. For this, two guiding questions, inclusion, exclusion and quality criteria were elaborated, in addition to a search in journal databases, in order to find the works that are more adherent to the proposed theme.*

Results: *After conducting the search with the exclusion, inclusion and quality criteria, three articles were identified that had good adherence to the proposed topic. The authors aimed to apply Machine Learning in the academic context for different purposes. All of them had good results, where it was possible to observe the application of the techniques for several purposes and how it can optimize school and academic management, bringing an optimization of time, both for teachers and students.*

Conclusion: *Through the study carried out, it is possible to conclude that Machine Learning and Data Processing has a great applicability in the field of education, but more research in the area is still needed.*

Key Word*: Education, Machine Learning, Data Processing, Educational Data Mining*

I.   Introduction

Signal processing is the field of science that studies methods of analysis and synthesis of signals, sequences or measurements that can represent physical quantities. In this method mathematical tools are used to describe the transformation of signals, tools such as linear algebra, statistics, complex and real analysis, etc. With the development of digital computers, processing methods have become a super efficient method, which has allowed their use in many areas of science and engineering. With this growth in the use of computer systems, many day-to-day activities today are performed by means of software developed for various purposes. With all this growth in technology, it is increasingly necessary to disseminate knowledge about Artificial Intelligence and Machine Learning, which are the most present areas in our society today due to technological advances.

This technological advance also drives a large growth in the volume and storage of educational and student data, and as these large amounts of information need to be analyzed, it becomes important to use computational resources to optimize time and quality. This opportunizes the use of Educational Data Mining

---

(EDM) to make the teaching and learning processes more efficient, since this methodology can be applied in several areas for various purposes (Vivian, 2022). Marques (2022), explains that in the EAW method there is the application of data mining techniques and can also have the application of probabilistic models, in data in the area of education. To apply the method it is important to understand which tools to use and also understand which techniques best fit the context studied. One solution used to know and understand the main tools used, techniques and factors explored in the context studied, is to perform a Systematic Literature Review (SLR), which is an analysis of materials found in the literature, which exposes the methods that can be replicated.

Through the context presented, the main objective of this article is to perform a systematic review of the literature in order to understand how the Signal Processing and Machine Learning is applied in the area of Education. This study aims to understand how the machine learning technique assists in educational management in better decision making. The justification of the study occurs due to the technological evolution, it must be better studied and understood in the educational area, so that it can provide and optimize the processes involved in this area.

II.   Materials and Methods

The article is a literature search on educational data processing techniques, for which a state of the art analysis was performed through a Systematic Literature Review. A total of 3 articles were used to perform the necessary analyses. Figure 1 describes the stages of the search process and selection of articles, along with the total number of articles found in the search, through the use of inclusion, exclusion and quality criteria.
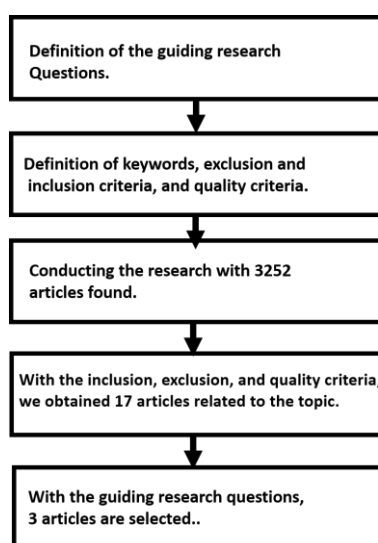
Table 1: Results of the surveys conducted.

Base	Phase 1	Phase 2	Phase 3
CAPES Periodicals	82	2	1
Scholar Google	3170	15	2
Total	3252	17	3

Inclusion Criteria:
1.   Articles, dissertations and journal articles;
2.   Publications that are within the proposed theme, and that are national and international productions, in Portuguese and English;
3.   Publications from the last 10 years;

Exclusion Criteria:
1.   Research papers and studies that are not related to the subject and that have been published more than 10 years ago;
2.   Papers that are not scientific articles, dissertations, or academic papers;
3.   Study published in a language other than Portuguese and English;

Quality Criteria:
1. Is there a clear definition of the research objectives in the article?
2. Is there a good description of the results found?
3. Does the study have theoretical value for the research conducted?

Methodological Procedure

During the development of the Systematic Literature Review, two questions were elaborated to guide the study and the analysis of the articles found. They are the following:

Q1: How is the machine learning technique used for educational purposes?

Q2: What is the authors' conclusion regarding the techniques used?

Search String

To carry out this review, key words were defined, and then a search was made in digital libraries, which will be used as a source to study the proposed theme.

Table 2: Search string used in the databases.

Operators	OR	AND
Search String		("machine learning" OR "machine learning" OR "deep learning" OR "deep learning" OR "neural networks" OR "neural networks" OR "learning algorithms") AND ("signal processing" OR "signal processing" OR "digital signal processing" OR "signal analysis" OR "image processing" OR "audio processing" OR "speech recognition") AND ("statistics" OR "statistics" OR "statistical analysis" OR "statistical modeling") AND ("education" OR "education" OR "educational" OR "educational technology" OR "e-learning" OR "e-learning" OR "e-learning" OR "educational assessment")

III. Result

After selecting the articles and using the inclusion, exclusion, and quality criteria, we came up with 3 articles that stood out within the theme addressed.

Table 3: Selected Articles.

Authors	Year of publication	Topic
Rolim, V., Ferreira, R. & Costa, E.	2017	Monitoring Educational Forums usingMachine Learning Techniques for Monitoring Educational Forums.
SOUZA, V. F. de; SANTOS, T. C. B. dos	2021	Educational Data Mining Process Applied to Predicting Student Performance: A Comparison of Machine Learning and Deep Learning Techniques
PINHO, C. M. de A. .; MOURA, A. F. de; GASPAR, M. A.; NAPOLITANO, D. M. R.	2022	Identifying deficiencies in educational texts with the application of natural language processing and machine learning.

Q1: How is the machine learning technique used for educational purposes?

Souza and Santos (2021) conducted a study where the objective was the prediction of student performance in a public dataset, comparing Machine Learning and Deep Learning techniques, indicating which are the main predictor attributes for the performance of these students. The authors implemented an educational data mining process for their study, which consisted of 4 stages, the first stage was data collection, the second stage consisted of feature extraction and data cleaning, the third stage was analytical processing and algorithm, and the fourth stage of the process was analysis and interpretation of the results

found. The result found by the authors showed that regarding the educational data mining techniques, the results are 66% accuracy for Naive Bayes, 87% for Decision Tree, 83% for Random Forest, 82% for Support Vector Machine and 94% for RNAM.

The correction of texts and discursive questions is very important in the academic environment, important tests such as ENEM, require a lot of demand and the Ministry of Education, in 2019, has already announced that there is a trend that in the future the test will be held digitally. Within this context, Pinho, et al. (2022), conducted a study whose main objective was to analyze texts using Machine Learning and Natural Language Processing techniques in order to identify deficiencies and potential errors in educational texts. To this end, the authors classified 695 essays written in Portuguese language into 20 themes. The authors' work consisted of 5 stages, the first was data acquisition and selection, the second was preprocessing, the third was data formatting and transformation, the fourth stage was data mining and visualized patterns, and the last stage consisted of interpretation and analysis of the results. To carry out the process proposed by the authors, the main language tool used was Python version 3.7. The Python libraries used were NLTK for text normalization, Numpy for linear algebra and matrix operations, pandas for data management, Scikit-learn for attribute extraction and machine learning algorithms, and Matplotlib for data visualization.

Distance education has grown a lot in Brazil, this modality allows students to have flexibility and practicality in their routine, however, one of the great difficulties is to follow so many students in the same class, especially when it comes to online forums and activities. Rolim, Ferreira and Costa (2017), conducted a study where the main objective of the authors was to help both teachers and students in the realization and monitoring of online forums. The authors performed for this the classification of posts and content recommendation. The authors performed for this the identification of doubt posts, extraction of the subject of the post and then evaluation in an experimental study. In the identification, the accuracy was 97% with the method used of neural network-based algorithm, the extraction step was performed using an unsupervised learning algorithm and showed an accuracy of 76.1%. With these implementations performed by the author, the system recognizes the student's doubt through the extraction of the message and recommends to him the most appropriate video.

Q2: What is the authors' conclusion regarding the techniques used?

Souza and Santos (2021) concluded that according to the accuracy found in models generated from Deep Learning, this technique has a higher performance when compared to Machine Learning. The authors also concluded that attributes related to school activities are more predictive of student performance than demographic and socioeconomic data.

Pinho, et al. (2022), were able, through the applied technique, to identify the essays whose theme runs away from what was proposed, in addition to other information that enables the teacher to identify possible errors and flaws in the writing of the text, such as insufficient text and textual cohesion. The authors, through their results, concluded that the AI techniques are a great way to support teachers, providing them with optimization of the correction time of tests and essays.

Rolim, Ferreira, and Costa (2017), had a very interesting idea and their technique showed good accuracy, the authors concluded that this technique helps not only teachers, in optimizing time in corrections and answers, but also students, since the system offers them the most appropriate content, which can solve their doubts.

IV. Conclusion

During the Systematic Literature Review performed, it was possible to realize that the Educational Data Mining is a very comprehensive area, but it still needs to be further studied and structured, since the data are diverse and very voluminous, since the educational area is very broad. Several authors conduct studies involving Machine Learning and Statistical Data Processing, but the area still needs much deepening and study, however, evaluating the results found by the authors, it was possible to realize that the results tend to be very positive and useful for education.

The articles selected through the inclusion, exclusion, and quality criteria presented good theoretical content, where it was possible to understand well the theme addressed, its importance, and how it can influence educational management.

It was possible through this research to understand that the Machine Learning and Statistical Processing method has several applicabilities in the area of education, several authors have made themes of great importance, both for students and teachers, and that it can be used to optimize the pedagogical process, making this interaction between student and teacher more practical and with faster returns.

References

[1]. VIVIAN, R. L. et al. Educational Data Mining and Sentiment Analysis in Virtual Learning Environments: a Systematic Mapping. EaD em Foco, v. 12, n. 2, e1786, 2022.

[2]. PINHO, C. M. de A. .; MOURA, A. F. de; GASPAR, M. A.; NAPOLITANO, D. M. R. Identification of deficiencies in educational texts with the application of natural language processing and machine learning. ETD - Educação Temática Digital, [S. l.], v. 24, n. 2, p. 350-372, 2022.

[3]. RIBEIRO, N.L., MOTA, J.C.M., ALMEIDA, A.L.F. Tensor Signal Processing Applied to Communications. JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY, VOL. 5, NO. 2, OCTOBER 2015.

[4]. ALMEIDA, A.L.F. Tensor Modeling and signal processing for wireless communication systems. Universidade Federal do Ceará, Fortaleza - Ceará, 2007.

[5]. SOUZA, V.F., CAZELLA, S.C. Mining educational data with regression algorithms: a study on performance prediction. Revista Educar Mais, [S. l.], v. 6, p. 183-198, 2022.

[6]. SOUZA, V.F., SANTOS, T.C.B. Processo de Mineração de Dados Educacionais aplicado na Predvisão do Desempenho de Alunos: Uma comparação entre as Técnicas de Aprendizagem de Máquina e Aprendizagem Profunda. Revista Brasileira de Informática na Educação - RBIE, 2021.

[7]. CHOJI, M.F; DAMASCENO, C. D. N; BITTENCOURT, I. I; ISOTANI, S. Data mining of the Enade from 2016 to 2018: an analysis on the municipality of Araçatuba/SP. RENOTE, Porto Alegre, v. 19, n. 2, p. 183-192, 2021.

[8]. FERNANDES, F. T.; CHIAVEGATTO, A. D. P. Perspectives of using data mining and machine learning in occupational health and safety. Brazilian Journal of Occupational Health. 2019, v. 44.

[9]. SOUZA, V. F. de; SANTOS, T. C. B. dos. Processo de Mineração de Dados Educacionais aplicado na Predação do Desempenho de Alunos: Uma comparação entre as Técnicas de Aprendizagem de Máquina e Aprendizagem Profunda. Revista Brasileira de Informática na Educação, [S. l.], v. 29, p. 519-546, 2021.

[10]. Rolim, V., Ferreira, R. & Costa, E. (2017). Monitoring Educational Forums using Machine Learning Techniques for Monitoring Educational Forums. Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação - RBIE), 25(3).

[11]. MARQUES, L.T., MARQUES, B.T., SILVA, C.A.M. . The Discovery of the Causes of Academic Retention Using Data Mining: A Systematic Literature Review. RENOTE, Porto Alegre, v. 20, n. 1, p. 263-272, 2022.