MACHINE LEARNING IN EDUCATION - A SURVEY OF CURRENT RESEARCH TRENDS

Danijel Kučak, Vedran Juričić, Goran Đambić

Abstract

Nowadays, Machine Learning (ML) is one of the most promising application areas in a field of Information Technology where its application scope is almost unlimited. The application of machine learning in an education area is currently very interesting to researchers and scientists, and it is the main focus of our study. The aim of this paper is to evaluate the possibilities of applying and using machine learning in the education area. This paper identifies and analyses suitable literature, research papers and articles in order to determine their categorization in the field of education, to determine the current trends of using machine learning in education, and to determine its current and future applications.

Keywords: machine learning; education; student performance; student retention

1. Introduction

The goal of machine learning is to program computers to use example data or past experience to solve a given problem [1]. With machine learning, for example, self-driving vehicles are really close to be on roads every day. Patter recognition, education, computer vision, bioinformatics, natural language processing, etc. are just some of fields where machine learning can be applied at.

This study is focused on application of machine learning in education. Education is changing on daily basis. There aren't just students in classroom anymore, looking at the notebook, while a teacher lectures. Today's classrooms use digital resources and are investing in machine learning. In education, for example, machine learning could be applied to support teachers, predict student performance, test students, etc.

The goal of our research was to categorize studies about the usage of machine learning in education area. Before research has been started, we choose the research questions:

- Research question 1: How to categorize research papers in the area of using machine learning in education?
- Research question 2: What are trends in application of the machine learning in education area?

A comprehensive literature study has been done to get answer on these questions. We used the Systematic Literature Review approach (SLR) [2] to collect primary studies regarding this research scope. The main aim was to classify studies and the selected studies were then analysed and categorized using the content analysis method. As result of the classification, four main categories were defined:

- Grading students
- Improving student retention
- Predicting student performance
- Testing students

The structure of this study is as follows: In the second section, we reviewed machine learning and its application in education area. In the third section, the research approach used in this study was introduced. After that, analysis of the results and categories of studies was introduced. Fifth section includes a conclusion and suggestions for future research on the topic.

2. Machine learning and its application in education area

Machine learning can be considered a part of artificial intelligence (AI). Machine learning is, at its core, the process of granting a machine or model access to data and letting it learn for itself. In 1959, Arthur Samuel came up with the brilliant idea that we should not have to teach computers, but rather, we could let them learn on their own. He coined the term "machine learning" to describe his theory, which is now a standard definition for the ability of computers to learn autonomously [3].

Machine learning is programming computers to optimize a performance criterion using example data or experience [1]. Implementing a machine learning algorithm means implementing a model that outputs correct information given that we have provided input data. You can think of a model as a black box: data go in at the beginning, and some other data go out at the end — but the processes in between are complex. For example, if we want to create a model that predicts what the house price in some region next year will be, based on situation on the market for the last three years, we would feed the model metrics such as house prices on the market in last three years, interest rates and salary rates. The output would be the house price prediction for the next year. The process through which a model learns how to make sense of input data is called "model training". Training is a key concept in machine learning.

The best way to describe the potential of machine learning is to explore how people and companies are currently taking advantage of it. Some examples could be:

- Natural language processing: - Google Translate is created from a set of machine learning algorithms that updates the service over time based on input from users, like new words and syntax. Siri, Alexa, Cortana, and, most recently, Google Assistant all rely on natural language processing to recognize speech and synthesis, allowing them to understand or pronounce words they have never encountered before.
- Recommendation systems: - On Netflix, Amazon, Google, etc., everything that is recommended to you depends on your search activity. These websites deliver recommendations across platforms, devices, and apps. Machines match buyers with sellers and their products, digital content with viewers who want to see them - all of which improves our online experiences significantly. Amazon has machine learning algorithms in place that it can predict with high certainty what you'll buy and when you'll buy it. The company even owns a patent for "anticipatory shipping," a system that ships a product to the nearest warehouse so you can order and receive your item on the same day (although it is unclear whether they have implemented it yet).
- Algorithmic trading: - Algorithmic trading is a process that involves random behaviour, ever-changing data, and a variety of factors. While financiers cannot predict all of that behaviour, machine learning algorithms can — and they respond to changes in the market much faster than a human.

There are plenty of other business implementations of machine learning, lot of them are in education area [4]. Some of interesting areas are:

- Predict Student Performance (A great application of machine learning is predicting student performance. By "learning" about each student, the machine learning model can find out weaknesses and suggests ways to improve, such as additional lectures or study additional literature [5].)
- Test Students & Grade Students Fairly (Machine learning can help creating computerized adaptive assessments [6]. The machine learning based assessment provides constant feedback to teachers and students about how the student learns, the support they need and the progress they are making towards their learning goals [7].)
- Improve Retention (Machine learning, such as learning analytics, will also help improve retention rates. By identifying "at risk" students, schools can reach out to those students and get them the help they need to be successful [8].)
- Support teachers and institution stuff (Machine learning based algorithms can help with classification of students handwritten assessment papers [21])

3. Research

The above-mentioned research questions were extracted when we studied the aims of this research. To answer these research questions, we performed a literature study. We decided to use the SLR method to collect the relevant primary studies and followed the guidelines given by [2].

For the SLR, we decided to do an electronic search. The databases used were IEEE Xplore Digital Library (IEEE), Scopus database, Web of Science database (WoS), ScienceDirect, and others (Google Scholar and similar). The search term was ("Machine Learning" AND "Education") OR ("Machine Learning" AND "Educational"). The targeted amount of related studies was between fifty and one hundred, because this amount would give us enough information for categorization and research trends.

In the next phase, we created four categories and classified the papers into those. The reason for this classification was that most of the papers published were relatively distinctive in terms of the research objective, methodology, and application. To be simplest as possible, we created four categories of studies without ignoring the variations of themes. This way, we examined the research papers that fell under the same. The final number of papers collected was 67. Some studies was included in more than just one category.

The research categories and the total number of selected studies are as follows:
- (A) Grading students (12 studies)
- (B) Improving student retention (17 studies)
- (C) Predicting student performance (42 studies)
- (D) Testing students (6 studies)

Letters A, B, C and D are just used as tags and serve as a mark for further referencing in the rest of this study.

4. Results

This section describes the main characteristics of each category. We present all four main categories in subsections, supported with relevant papers.

4.1. Grading students

Machine learning can grade students by removing human biases. Some recent examples are use of the supervised Machine Learning for text classification to predict students' final course grades in some course and exhibited the potential of using ML classified messages to identify students at risk of course failure [9]. In addition, there are aims to improve the assessment of problem solving in education by employing language technologies and computational-statistical machine learning methods to grade students' natural language responses automatically [10]. Great example of use machine learning for grading students is by comparing their actions to a model of expert behaviour [11].

4.2. Improving student retention

As we said before, by identifying "at risk" students early, schools can detect and contact those students and help them to be more successful. Student retention is an essential part of many enrolment systems. It affects almost all segments of university or school metrics: reputation, financials, ranking. Specially, student retention has become one of the most important things for managers in higher education institutions. There are few studies, which developed models to predict and to introduce the reasons behind student's number decreasing [8] [12] [13] [14] [15] [16].

4.3. Predicting student performance

Probably a major benefit of machine learning (regarding number of studies in scientific databases) is its ability to predict student performance. By "learning" about each student, the technology can identify weaknesses and suggests ways to improve, such as additional practice tests. This seems to be very hot research trend; there are lot of studies in recent years in this area, as we said before. For example, study [17] employs the machine learning approach called the Recursive Clustering technique to group the students of the programming course into groups based on their performance in the prerequisite courses, co-requisite and current course work result. Students present in the lower groups will be taken into consideration since they are highly prone to fail. In another interesting study in this category, authors have proposed a new model to categorize students into three categories to determine their learning capabilities and to help them to improve their studying techniques. They have chosen the state of the art of machine learning approach to classify student's nature of study by selecting prominent features of their activity in their academic field. They have chosen a data driven approach where key factors that determines the base of student and classify them into high, medium and low ranks [18]. Yet another study in this category, but from other perspective brings [19].

4.4 . Testing students

The machine learning based assessment provides constant feedback to teachers, students and parents about how the student learns, the support they need and the progress they are making towards their learning goals. Authors in study [20] introduced a system for training of students' ability to construct correct proofs in propositional or predicate logic. In addition to common techniques, including presentations supported by slides and exercises they used animations, which were based on carefully selected demonstrative examples and their step-by step solutions. In order to test students' knowledge, they prepared a questionnaire that captured the entire process of a logic proof construction. A student constructed a proof and then answered questions from the questionnaire. They described the design of the questionnaire and discussed its dis/advantages. At the end, they then applied frequent subgraph mining together with supervised machine learning algorithms to perform an automatic evaluation of correctness of the proofs.

5. Conclusion

The aim of this study was to evaluate the current state of the art in the application of machine learning in education area. The amount of studies (papers and articles) was large, so only some of studies, which we found as good representatives, were mentioned in results this study. This study shows that there are significant different ways to benefit from machine learning application in education area.

As we stated in introduction section, one of our goals was try to classify studies in the field of machine learning application in education area. Based on our survey, the papers reviewed under category marked as A research a ways how machine learning can grade students by removing human biases (fairly grading)

Reviewing studies under category marked as B, showed how machine-learning algorithms can help schools or faculties to reach out to students and get them the help they need to be successful as early as possible. Student retention is an essential part of many enrolment management systems. It affects university rankings, school reputation, and financial wellbeing. Student retention has become one of the most important priorities for decision makers in higher education institutions, so there are lot of studies in that category.

Reviewing studies under category marked as C, showed us how major benefit of machine learning (regarding number of studies in scientific databases) is its ability to predict student performance. By "learning" about each student, the technology can identify weaknesses and suggests ways to improve. According to our survey, this is most interesting area of machine learning application to researchers. There are lot of studies in recent years in that category, and lot of machine learning models were provided to predict student performance on different parameters. We would say that this category is definitely the trend.

Reviewing studies under category marked as D, showed some models how machine learning can help move away from standardized testing. Machine learning based assessment provides constant feedback to teachers, students and parents about how the student learns, the support they need and the progress they are making towards their learning goals.

As we have found earlier, research has been made over several relevant databases, but of course, not all were involved, so this can be considered as limitation of study. In addition, there is a possibility that some of the relevant studies may be skipped by chance.

In the future, we plan to implement own machine learning model for suggesting potential student to enrol or not to enrol on University College algebra, Study of Software Engineering, based on different parameters. As we have rich database with lot of information of students on previous years, we believe that study would be of help to support our admission office as help in student enrolment process.

6. References

[1] Alpaydin, E. (2009). Introduction to machine learning. MIT press.
[2] Kitchenham. B.A., Charters,S. (2007) Guidelines for performing Systematic Literature Reviews in Software Engineering. Version 2.3, EBSE Technical Report EBSE-2007-01, Keele University, Keele, Staffordshire, United Kingdom
[3] Samuel, A.L. (1959) Some Studies in Machine Learning Using the Game of Checkers, IBM Journal of Research and Development, vol. 3, no. 3, pp. 210-229, doi: 10.1147/rd.33.0210
[4] https://www.thetechedvocate.org/8-ways-machine-learning-will-improve-education, (2018), Accessed on: 2018-09-25
[5] Anozie, N., Junker, B. W. (2006, July). Predicting end-of-year accountability assessment scores from monthly student records in an online tutoring system. Educational Data Mining: Papers from the AAAI Workshop. Menlo Park, CA: AAAI Press.
[6] Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education.
[7] El-Alfy, E. S. M., & Abdel-Aal, R. E. (2008). Construction and analysis of educational tests using abductive machine learning. Computers & Education, 51(1), 1-16.

[8]   Đambić, G., Krajcar, M. & Bele, D. (2016). Machine learning model for early detection of higher education students that need additional attention in introductory programming courses. International Journal of Digital Technology & Economy, 1 (1), 1-11

[9]   Wu, J. Y., Hsiao, Y. C., & Nian, M. W. (2018). Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. Interactive Learning Environments, 1-16.

[10] Wang, H. C., Chang, C. Y., & Li, T. Y. (2008). Assessing creative problem-solving with automated text grading. Computers & Education, 51(4), 1450-1466.

[11] Abbott, R. G. (2006, June). Automated expert modeling for automated student evaluation. In International Conference on Intelligent Tutoring Systems (pp. 1-10). Springer, Berlin, Heidelberg.

[12] Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498-506.

[13] Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. Computers & Education, 53(3), 950-965.

[14] Ram, S., Wang, Y., Currim, F., & Currim, S. (2015). Using big data for predicting freshmen retention.

[15] Chai, K. E., & Gibson, D. (2015). Predicting the Risk of Attrition for Undergraduate Students with Time Based Modelling. International Association for Development of the Information Society.

[16] Jia, J. W., & Mareboyana, M. (2014). Predictive models for undergraduate student retention using machine learning algorithms. In Transactions on Engineering Technologies (pp. 315-329). Springer, Dordrecht.

[17] Anand, V. K., Rahiman, S. A., George, E. B., & Huda, A. S. (2018, March). Recursive clustering technique for students' performance evaluation in programming courses. In Majan International Conference (MIC), 2018 (pp. 1-5). IEEE.

[18] Alam, M. M., Mohiuddin, K., Das, A. K., Islam, M. K., Kaonain, M. S., & Ali, M. H. (2018, March). A Reduced feature based neural network approach to classify the category of students. In Proceedings of the 2nd International Conference on Innovation in Artificial Intelligence (pp. 28-32). ACM.

[19] Ciolacu, M., Tehrani, A. F., Beer, R., & Popp, H. (2017, October). Education 4.0—Fostering student's performance with machine learning methods. In Design and Technology in Electronic Packaging (SIITME), 2017 IEEE 23rd International Symposium for (pp. 438-443). IEEE.

[20] Vaculík, K., Popelínský, L., Mráková, E., & Jurco, J. (2013, October). Tutoring and automatic evaluation of logic proofs. In European Conference on e-Learning (p. 495). Academic Conferences International Limited.

[21] Celar, S., Stojkic, Z., Seremet, Z., Marusic, Z., & Zelenika, D. (2015). Classification of Test Documents Based on Handwritten Student ID's Characteristics. Annals of DAAAM and Proceedings of DAAAM Symposium. 2014. ISBN 978-3-901509-99-5., 782-790.