Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review

Lidia Sandra [1], Ford Lumbangaol [2], Tokuro Matsuo [3]

[1]*Pscyholgy Department, Krida Wacana Christian University, Jakarta, Indonesia*
[2]*Computer Science Department,Bina Nusantara University, Jakarta, Indonesia*
[3]*Advanced Institute of Industrial Technology, Tokyo, Japan*

Abstract **- One of the ultimate goals of the learning process is the success of student learning. Using data and students' achievement with machine learning to predict the success of student learning will be a crucial contribution to everyone involved in determining appropriate strategies to help students perform. The selected 11 research articles were chosen using the inclusion criteria from 2753 articles from the IEEE Access and Science Direct database that was dated within 2019-2021 and 285 articles that were research articles. This study found that the classification machine learning algorithm was most often used in predicting the success of students' learning. Four algorithms that were used most often to predict the success of students' learning are ANN, Naïve Bayes, Logistic Regression, SVM and Decision Tree. Meanwhile, the data used in these research articles predominantly classified students' success in learning into two or three categories which are pass/fail; or fail/pass/excellent.**

Keywords **- machine learning algorithm, systematic literature review, student's performance.**

1. Introduction

The rapid digitization of educational institutions is very significant in helping educators and education personnel to collect data. The abundance of data on educational institutions is also moving very dynamically accompanied by changes in learning patterns and governance of educational institutions from previously offline-based to online-based. Similar to other industries affected by the pattern of 4.0 Industry, explicit programming which is currently commonly used as a tool to provide solutions to problems is no longer sufficient in some cases. In this case, machine learning fills the gap and provides a solution for processing these very large and dynamic data.

Prediction of student's performance can be used as a basis for early intervention on the potential failure of students to achieve learning objectives; and at the same time able to make changes to learning strategies in order to facilitate student diversity [1]. This is also supported by the availability of student data that can be processed to make predictions such as behavioral data, type and frequency of activities carried out (both in online and offline learning settings), age, height and weight, previous academic achievement; as well as latent data such as personality and motivation as well as external data such as parenting patterns, parental support makes the data contained in educational institutions abundant, multi-data and qualified to be analyzed.

This study aims to systematically examine the use of machine learning algorithms to predict student achievement. As of May 2021, no similar research has been found based on searches on the IEEE Access and Science Direct databases. So that this article is expected to make a scientific contribution to researchers in the area of predicting student learning achievement, computer science, and statistics (or data science). It is hoped that the results of this systematic literature review can provide insight into the most effective machine learning algorithms for predicting student achievement and increase stakeholders' understanding of the implementation of machine learning to predict student achievement.

DOI: 10.18421/TEM104-56
https://doi.org/10.18421/TEM104-56

Corresponding author: Lidia Sandra,
Pscyholgy Department, Krida Wacana Christian University, Jakarta, Indonesia.
Email: lidia.sandra@ukrida.ac.id

Received: 03 August 2021.
Revised: 11 November 2021.
Accepted: 18 November 2021.
Published: 26 November 2021.

A. Student Performance

Student's performance in the educational process can literally be defined as something that is obtained from changes in the behavior of students based on their experiences, besides that learning outcomes are also a realization of the potential or capacity possessed by students [1]. These learning outcomes from students can be seen from their behavior, both behavior in the form of understanding knowledge, thinking skills, or motor skills [2]; an outcome of the process of changing student behavior after attending lessons [3]. The concrete form of student's performance can be seen from their understanding of the knowledge being studied, their expertise in processing information and making decisions based on certain thoughts or motor skills [4]. Based on those understandings, student's performance can be observed and measured in the realm of students' knowledge, attitudes and skills after following a series of lessons. Student's performance depends on the teaching and learning process they go through, so that learning outcomes can be used as considerations in improving the quality of the learning process.

On the other hand, a similar terminology is learning achievement which is a measure of student achievement after participating in learning activities in the form of an assessment scale (either letters, numbers or certain symbols [5]. Giving a weight or rating for student's achievement in learning requires the preparation of suitable assessment indicators, and needs to be ensured of the validity and the reliability [6]. This value can then be used as a description of student's performance in a certain period of time [7]. Student's performance is obtained after passing various measurements (in various forms of assessment) after students carry out several learning processes.

B. Machine Learning

Machine learning is categorized as a sub-section of artificial intelligence. In principle, what machine learning does is build mathematical models to help make sense of the data. This mathematical model consists of various variables that can change, so that the program formed with the mathematical model allows the program to adapt to changes in data [8]. Machine learning algorithms can be divided into two types, namely supervised learning and unsupervised learning. Supervised learning is a category in machine learning that is done by labeling the expected input and output. Algorithms in supervised learning will then produce output without human intervention such as programming modifications. One example of the use of supervised learning algorithms is the separation of spam and non-spam in email based on a mathematical model that was built previously using spam and non-spam email labels [9].

Supervised learning is divided into two types, namely classification and regression. The purpose of classification is to predict the class in the labeled data, which has provided a choice from a list of possibilities. Meanwhile, the main purpose of regression is to predict the description of the regression relationship from the data, namely a floating-point number.

The second type is unsupervised learning. In unsupervised learning, labeling is not done on all input data. Algorithms in supervised learning will read the data as it is without any relation to the label given so that the program will draw information directly from the existing data [8]. In general, there are two things that are generally done in unsupervised learning, namely by transforming datasets or by clustering. Dataset transformation is carried out with an algorithm that allows the program to convert existing datasets into datasets that are easier to understand [9].

While in clustering, the algorithm will separate the data into several clusters based on the similarity of data types [9]. One example of the application of the clustering algorithm is the compilation of photos uploaded to social media into albums based on the faces of those who appear in the photos. Solving problems using machine learning effectively requires an understanding of the problem to be solved in its entirety. Understanding the problem will affect the decision on what data is needed and what algorithm should be used to solve the problem [8]. Algorithms that will be used in machine learning (both categorized in supervised learning and unsupervised learning) need to be adapted to the existing data and the analysis objectives to be achieved [9].

2. Method

This study uses a systematic literature review method which is preceded by making a systematic literature review protocol. This method was chosen based on the consideration of the lack of systematic literature review on the implementation of machine learning algorithms to predict student's performance until May 2021 on the IEEE Access and Science Direct databases. This systematic literature review is expected to be a stimulus for further data science research that contributes to the field of education as quality education is one of the important issues targeted in the Sustainable Development Goals action plan [10].

A. Systematic Literature Review

Systematic literature review is a literature review method that aims to answer research questions by identifying, assessing, evaluating and interpreting all findings related to the research topic. Systematic literature review has proven to be an effective research method to provide an overview of trends in certain research topics, both results, methodology and coverage of previous research fields [11].

The systematic literature review method is carried out by following certain protocols. The stages of implementing a systematic literature review begin with formulating research questions and making a systematic literature review protocol. The literature review protocol in question is to determine what keywords to use in searching for articles in the database and what criteria to cover (inclusion criteria) and what criteria for articles that are not used (exclusion criteria). After the articles through the protocol are collected, the next step is to identify the appropriate literature, select from the primary study, extract data and assess the quality of the research results obtained. In the final stage, a synthesis of various research results is carried out and the stage of writing a systematic literature review article can be carried out (see Figure 1.) [11].

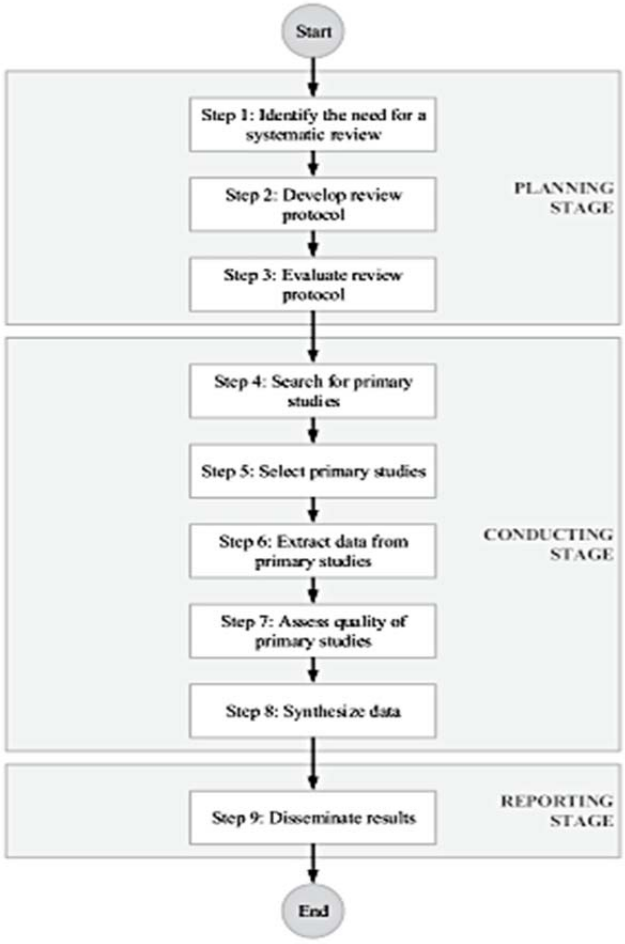

Figure 1. Systematic literature review stages [11]

B. Research Questions

The main purpose of this systematic literature review is to explore the use of machine learning algorithms in predicting student's performance. To achieve this goal, the research questions (RQ) that are prepared are:

RQ1: What types of machine learning algorithms are used to predict student's performance?
RQ2: How are predictions of students' performance made by machine learning algorithms useful?
RQ3: What are the machine learning algorithms or techniques implemented to predict student's performance?
RQ4: What are the results and conclusions obtained from the implementation of machine learning in predicting student's performance?

C. Data Collection

The study was conducted in mid-May 2021. The standard rules in searching the database were carried out according to the exclusion and inclusion criteria, as well as the search results. The data collected is sourced from the IEEE Access and Science Direct databases. While the keywords used in the search are "student's performance OR students' performance" AND "machine learning algorithm" which is filled in the advanced search column.

The search is then selected based on the year of publication, the type of article and the abstract. The articles used as references are research articles published in the 2019-2021 range and written in English. The exclusion criteria are publications that are not available in full text, not in English and research that does not discuss predictions of students' performance and students' achievement based on machine learning algorithms. Student's performance covered is students' performance for all levels of education from basic education to higher education.

D. Data Analysis

To analyze the data, each research article selected as a sample in a systematic literature review is grouped based on research questions.

3. Result

A. Selected Articles

This research was conducted in early May 2021. In a search using the keywords "student's performance OR students' performance OR student's achievement OR students' achievement" AND "machine learning algorithm" published in 2019-2021, found 2,753 articles (2,694 articles in the database IEEE Access and 59 articles in the Science Direct database.) After

the selection was made based on scientific journal articles only, by exclusion of publication of articles in books, magazines and proceedings and selecting only articles in journals directly related to education and machine learning and selecting articles with open access to full text, we obtained 291 research articles (250 articles in the IEEE Access database and 41 articles in the Science Direct database).

B. Selection of Articles with Inclusion Criteria

The selection process started from 291 research articles. The title and abstract are read carefully to determine which articles are appropriate and which do not meet the inclusion criteria. After reading the title and abstract, there were 285 research articles that met the criteria.

After reading the title and abstract, the entire text of the article was also read to ensure its suitability. At that stage, there were 11 scientific articles that met the inclusion criteria; while the rest are not used because they do not meet the inclusion criteria, such as not predicting student's performance or not using machine learning algorithms. This sample of 11 articles discusses the implementation of machine learning algorithms in predicting student's performance (5 articles in the IEEE Access database and 6 articles in the Science Direct database), while the rest fall into the exclusion criteria. These 11 scientific articles are used as references in the preparation of a systematic literature review. After conducting a systematic literature review of 11 research articles, then the answers to the research questions that have been previously proposed are compiled.

C. Research Articles Used

A systematic literature review of 11 samples of research articles was obtained from 6 journals. The journals that were sampled the most were the IEEE Access journal and the Computer & Education Journal (see Figure 2.).

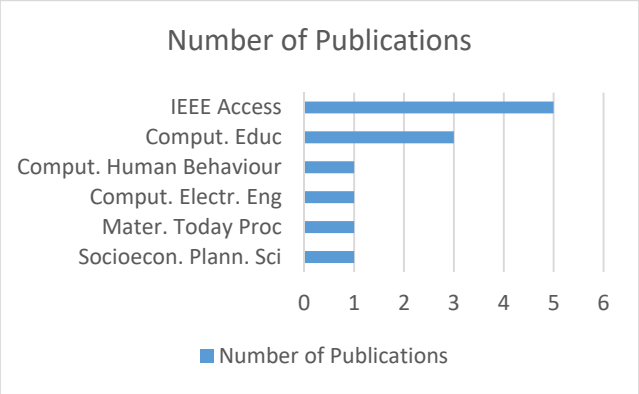

Figure 2. Number of publications – journal based

The most widely used journals are the IEEE Access journal and the Computer & Education journal (see Table 1.).

Table 1. Number of articles used based on the journal

No.	Journal	Number of Articles
1.	IEEE Access	5
2.	Computers in Human Behavior	1
3.	Computers & Education	3
4.	Computers & Electrical Engineering	1
5.	Socio-economic Planning Sciences	1
6.	Materials Today: Proceedings	1

The selected journals are international scientific journals and most of them are in the Q1 category (in the Scimago Journal Rank ranking) (see Table 2.).

Table 2. Scimago journal rank of the used journals

No.	Journal	SJR	Q Category	Reference
1.	IEEE Access	0.78	Q1 in Computer Science	[26]
2.	Computers in Human Behavior	2.17	Q1 in Arts & Humanities	[27]
3.	Computers & Education	3.05	Q1 in Computer Science	[28]
4.	Computers & Electrical Engineering	0.58	Q1 in Computer Science	[29]
5.	Socio-economic Planning Sciences	1.32	Q1 in Economics and Econometrics	[30]
6.	Materials Today: Proceedings	0.3	*Not Yet Assigned*	[31]

It can be seen that research on the use of machine learning algorithms to predict student's performance is relatively growing. This is indicated by the increasing number of studies conducted in the study area (see Figure 3.). At least the paper in 2021 because 2021 is not over yet. There will be a significant increase in 2020.

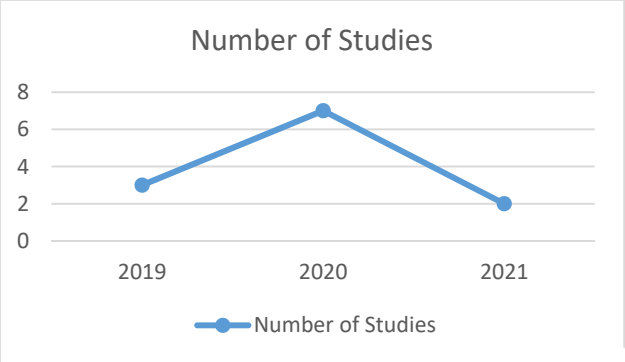

Figure 3. Number of studies – yearly based

D. Discussion of the Findings

The results of research from 11 scientific articles show that the type of machine learning algorithm used to predict student's performance, the algorithm found to be the most suitable in predicting the greatest student's performance is classification as can be seen in the image below (see Figure 4.).

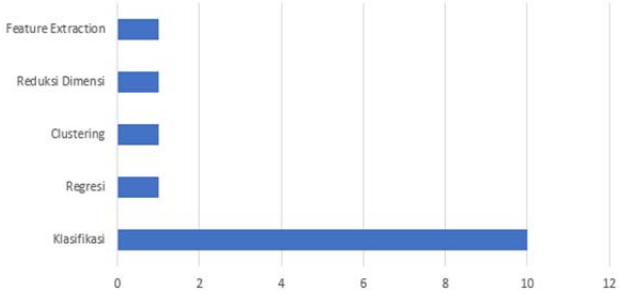

Figure 4. Machine learning algorithm usage types

In answering research question 2, how are predictions of students' performance made by machine learning algorithms useful? The results of this systematic literature review show that the prediction is useful in assisting teachers to predict student achievement earlier and in providing support for decision making to improve student's performance [12]. Machine learning algorithms are also relevant to be used to extract the criteria for student's performance in various educational domains to be able to create a coherent taxonomy of student learning outcomes [13]. In addition, machine learning algorithms can also be used to predict students who are likely to excel academically [14] and also detect students who are at risk of failure [15], [16]; so that students identified as being at risk of failure can be provided with additional support early and on time.

The increasingly massive learning process using the internet network on the one hand also makes the use of machine learning algorithms more relevant. The relationship between Internet usage behavior and academic performance can be used to predict students' academic performance [17]. In the context of using a blended learning platform, machine learning algorithms are used to model the predictions of academic performance of students using multiple-source and multimodal data [18] and predict student learning outcomes using data on student behavior and the training features used in online platforms [19].

The success of student learning can also be influenced by various factors such as the learning environment or other factors (either inside or outside of him). Machine learning algorithms were found to be relevant in exploring student learning behavior, solving students' academic problems, optimizing the educational environment, and enabling data-driven decision making [20]. Moreover, machine learning algorithms are also used to identify key factors that influence students' academic success in schools and explore the relationship between these key factors [21], [22]. For example, predicting student learning outcomes by combining various aspects of student life, namely student personality; behavior and learning styles as well as lifestyles such as sleep patterns, exercise patterns and others [23].

Machine learning algorithms can also be used to find the best method for resampling and classifying student learning outcomes datasets [24] and predicting student learning outcomes from achieving personalized assistance from family and school guidance [25]. Machine learning algorithms implemented in predicting student's performance include J48, JRIP, REPTree, Nnge, random tree, nave bayes, SMP, KNN, regression tree, random forest, decision tree, logistic regression, multilayer perceptron (MPL) neural network, support vector machine (SVM) and Artificial Neural Network (see Table 3.).

Table 3. Machine learning algorithms used

Algorithm	Reference
JRIP	[18]
Nnge	[18]
Naïve Bayes	[1], [16], [14], [23]
K-Nearest Neighbor (KNN)	[18], [24]
Regression tree	[21]
Random forest	[21], [24]
Decision tree (J48, RepTree, RandomTree, CART Decision Tree included)	[14], [15], [16], [17], [18], [1], [24]
PART	[18]
Long Short-Term Memory (LSTM)	[19], [22]
Logistic regression	[21], [18], [20], [12], [24], [14]
Support Vector Machine (SVM)	[14], [15], [18], [20], [19], [24]
Multilayer perceptron (MLP) *neural network*	[12], [16], [21]
Artificial Neural Network (ANN)	[16], [17], [15],

Algorithm	Reference
	[22], [14], [24]
Principal Component Analysis (PCA)	[14]
Feature Agglomeration (FA)	[14]
k-Means	[14]
Singular Value Decomposition (SVD)	[19]
XG- boost	[23]
ICGAN-DVSM	[24]
Recurrent neural network (RNN)	[19]
Bayesian knowledge tracing (BKT)	[19]
Deep knowledge tracing (DKT)	[19]

Various machine learning algorithms used to predict student's performance were tested for accuracy according to the data analysis stages required in machine learning algorithms and it was found that the most widely used machine learning algorithms to predict student's performance were Artificial Neural Network (ANN) algorithms, Support Vector Machine (SVM), Logistic Regression, and Decision Tree (See Figure 5.).

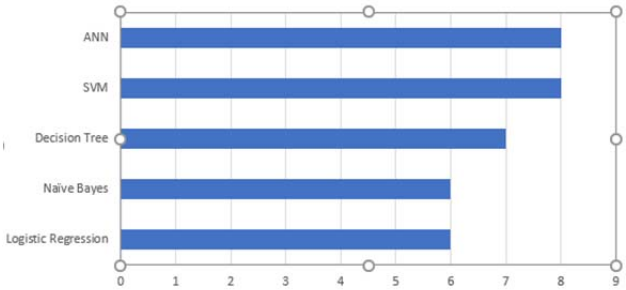

Figure 5. Machine learning algorithm used

While the results and conclusions obtained from the selected articles indicate that the implementation of machine learning to predict student's performance is as follows.

Table 4. Data, results and conclusions of machine learning algorithm implementation

Data, Results and Research Conclusions	
Student data at registration is used to predict student's performance in the form of student CGPA classification as follows: - Excellent (≥4.5) - Very good (3.75 to < 4.5) - Good (2.75 to < 3.75) - Average (2.0 to < 2.75) - Poor (<2.0) Algorithm used: ANN, Decision Tree, SVM and Naïve Bayes. The algorithm with the highest accuracy and precision is ANN; while the highest recall and F-1 measure is the DT algorithm.	14
Students' past learning achievement data (previous test scores), engagement (number of student visits to learning materials, search activities, participation in discussions), student demographics (age, expertise, date and time, location) are used to predict future test scores or assessments. Forward in the form of classification (PASS/ FAIL) and regression (range 0-100). Algorithm used: KNN, SVM, ANN, DT, Naïve Bayes, and Logistic Regression. The ANN algorithm has the highest level of accuracy when using a dataset of student engagement and learning success in the past.	16
Video data recorded from face-to-face classes (how many times students look at the computer, how many times they take notes in books, how many times they pay attention to the teacher), grades and the number of participations in practical classes, attendance and grades in online classes via the Moodle platform, and final exam scores to predict student's performance in the form of FAIL/ PASS/ DROPOUT classification. Algorithm used: J48, REPTREE, RamdomTree, JRIP, Nnge, PART. The Nnge and PART algorithms have the highest accuracy values in processing numerical and discretized data. It was found that the level of attention in theory class, quiz scores in Moodle, and activity in Moodle discussion forums were the most influential attributes in predicting final performance.	18
Learning activity data (how many times students do practice questions, scores obtained from practice questions, types of exercises done and duration of practice questions) are used to predict student's performance (range 0-100), using certain algorithms and frameworks. Algorithm used: DKT, RNN, BKT, and LSTM. However, this research does not focus on testing algorithms, but on testing frameworks. The frameworks tested are: MFA-DKT, BKT, DKT, DKT+, IRF-DKT and DKVMN. Compared to other frameworks, MFA-DKT is more significant in increasing the accuracy of predicting student's performance and being able to recommend appropriate learning programs. This is because MFA-DKT is able to combine learning activity data by using a Neural Network to distinguish one student's learning achievement from another.	19
Clickstream behavior data and assessment scores are used to predict student's performance which is classified into pass-fail/withdrawn-pass/distinction-fail/distinction-pass. Algorithm used: SVD, ANN, and logistic regression. The ANN algorithm is more accurate in predicting than other algorithms used.	
Data on the results of the 2016 SABER PRO test from 200,000 participants containing socioeconomic information, characteristics of high school and work status; including gender, age, and academic program undertaken. used to predict student's performance in college in the form of classification. Classification of student's performance rates: level 1 (0-125) / level 2 (126-	20

Data, Results and Research Conclusions	
155) / level 3 (156-200) / level 4 (201-300). Algorithm used: ANN and MLP. The ANN algorithm provides prediction results with the best level of accuracy.	
The data used is multi-source behavioral data inside and outside the classroom. Personality data (neuroticism, extraversion, and agreeableness), individual profiles (age, gender, height, weight, physical health, cardiorespiratory health, aerobic health, stress level), lifestyle data (diet, physical activity, time management) is used to predict student's performance in the form of classification: LOW/ MEDIUM/ HIGH PERFORMANCE. Algorithm used: LSTM. AugmentED model using LSTM algorithm can predict student's performance.	21
Data on gender, age, parents' health, education and occupation data, internet access, location of residence, previous scores were used to predict student's performance in 4 classifications (POOR, MEDIUM, GOOD, EXCELLENT). The algorithm used for the comparison of results: Random Forest, KNN, ANN, XG-boost, SVM, Decision Tree, Logistic Regression and Naïve Bayes. The Random Forest algorithm provides better prediction results than other algorithms used.	23
Data consisting of residence status (whether the student lives with parents or not), mother's education, mother's occupation, father's education, father's education, guardians of students, quality of relationships in the family, school support for education, and family support for education were used to predict success. student learning in the form of classification (PASS / FAIL). Algorithm used: ICGAN-DSVM. The ICGAN-DSVM algorithm has high accuracy in predicting student's performance.	24
Data on student interest, parental income, parental background, teacher involvement, friend involvement, social behavior and life style, job opportunities, awareness of education are used to predict student's performance whether students will succeed or fail (in the FAIL/PASS classification) using the KNN algorithm. The KNN algorithm is able to predict with a high degree of accuracy.	25

The results of research conducted in 11 research articles using various data available in educational institutions. Dominantly, predictions of student's performance are given in the form of classifications such as pass/fail; or file/pass/excellent. Meanwhile, clustering is done to find out what composition of student study groups will give the best learning achievement.

4. Conclusion

This Systematic Literature Review research aims to systematically examine the extent to which the implementation of machine learning algorithms and modeling has been carried out in predicting student's performance in the 2019-2021 range. Based on the exclusion and inclusion criteria, 11 research articles were used for review. This literature review is carried out by means of a systematic literature review, which is a technique to identify, collect and obtain information based on research questions that have been prepared previously. Four research questions in this study have been answered. The results showed that the most widely used type of machine learning algorithm to predict student's performance was the algorithm.

Predictions of students' performance made by machine learning algorithms are useful for helping teachers to predict student achievement earlier and providing support for decision making to improve student's performance, extracting student's performance criteria in various educational domains to be able to create a coherent taxonomy of student learning outcomes. In addition, machine learning algorithms can also be used to predict students who have a high chance of achieving academically and also detect students who have a risk of failure so that students who are identified as having a risk of failure can be given additional support early and on time. Machine learning algorithms are also used to explore student learning behavior, solve student academic problems, optimize the educational environment, and enable data-driven decision making, identify key factors that influence student academic success in schools and explore the relationships between these key factors and seek the best method for resampling and classifying student learning outcomes datasets and predicting student learning outcomes is the achievement of personalized assistance from family and school guidance.

The most widely used machine learning algorithms to predict student's performance are Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression, and Decision Tree algorithms.

The data used in this research article uses a variety of datasets. In general, predictions of student's performance are classified into 2 to 3 classifications such as pass/fail; or file/pass/excellent. Meanwhile, clustering is done to find out the composition of student learning groups which is how to best support student's performance.

Based on a systematic literature review of 11 research articles, it was found that relatively few similar studies have been carried out and are still growing. The use of LMS and internet networks in the learning process opens a huge space for research

that is oriented towards predicting student's performance, because the available data is getting bigger and more representative for students.

The challenge faced in using machine learning algorithms to predict student's performance is the difference in indicators for assessing student's performance in each class, subject and educational institution. In addition, the richness of various datasets that are processed in various types also provide additional challenges but also opportunities at the same time for the development of the use of machine learning algorithms to predict student's performance.

The findings from this systematic literature are expected to contribute to understanding for educational institutions to be able to use the most appropriate machine learning algorithm based on their needs and datasets to ensure the achievement of the ultimate learning goal, namely the learning success of students. Further research on the diversity of datasets used to predict student learning achievement and the development of machine learning algorithms and models that are able to facilitate the diversity of datasets to predict student learning achievement will be very useful for further development.

References

[1]. Sudjana, N. (2010). Penilaian Hasil Proses Belajar Mengajar.(Cet. XV). *Bandung: PT. Remaja Rosdakarya*.

[2]. Sukmadinata, N. S. (2009). Landasan Psikologi Proses Pendidikan: Bandung. *Remaja Rosdakarya.*

[3]. Jihad, A. (2008). *Evaluasi pembelajaran*. Multi Pressindo. Retrieved from: https://onesearch.id/Record/IOS3539.slims-374 [accessed: 12. June 2021].

[4]. Hamalik, O. (2001). Proses belajar mengajar. Retrieved from: https://onesearch.id/Record/IOS4317.laser-120011916 [accessed: 12. June 2021].

[5]. Dimyati & Mudjiyono.(2009). Belajar dan Pembelajaran. Jakarta: Rineka Cipta.

[6]. Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research papers in education*, *20*(3), 245-270. doi: 10.1080/02671520500193744

[7]. S. Suryabrata. (2011). Psikologi Pendidikan. Jakarta: Raja Grafindo Perkasa.

[8]. VanderPlas, J. (2017). Python Data Science Handbook. OReilly Media.

[9]. Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc."

[10]. Kementrian PPN. (2020). Pedoman Teknis Penyusunan Rencana Aksi Sustainable Development Goals, Jakarta: Kemnterian Perencanaan Pembangunana Nasional. (in Indonesian).

[11]. Wahono, R. S. (2015). A systematic literature review of software defect prediction. *Journal of Software Engineering*, *1*(1), 1-16.

[12]. Qu, S., Li, K., Zhang, S., & Wang, Y. (2018). Predicting achievement of students in smart campus. *IEEE Access*, *6*, 60264-60273. doi: 10.1109/ACCESS.2018.2875742.

[13]. Zughoul, O., Momani, F., Almasri, O. H., Zaidan, A. A., Zaidan, B. B., Alsalem, M. A., ... & Hashim, M. (2018). Comprehensive insights into the criteria of student performance in various educational domains. *IEEE access*, *6*, 73245-73264. doi: 10.1109/ACCESS.2018.2881282.

[14]. Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access*, *8*, 55462-55470. doi: 10.1109/ACCESS.2020.2981905.

[15]. Riestra-González, M., del Puerto Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*, *163*, 104108. doi: 10.1016/j.compedu.2020.104108

[16]. Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & education*, *143*, 103676. doi: 10.1016/j.compedu.2019.103676.

[17]. Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, *98*, 166-173. doi: 10.1016/j.chb.2019.04.015.

[18]. Chango, W., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Computers & Electrical Engineering*, *89*, 106908. doi: 10.1016/j.compeleceng.2020.106908.

[19]. Liu, D., Zhang, Y., Zhang, J., Li, Q., Zhang, C., & Yin, Y. (2020). Multiple Features Fusion Attention Mechanism Enhanced Deep Knowledge Tracing for Student Performance Prediction. *IEEE Access*, *8*, 194894-194903. doi: 10.1109/ACCESS.2020.3033200.

[20]. Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human behavior*, *104*, 106189. doi: 10.1016/j.chb.2019.106189

[21]. Rebai, S., Yahia, F. B., & Essid, H. (2020). A graphically based machine learning approach to predict secondary schools performance in Tunisia. *Socio-Economic Planning Sciences*, *70*, 100724. doi: 10.1016/j.seps.2019.06.009

[22]. Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, *2*, 100018. doi: 10.1016/j.caeai.2021.100018.

[23]. L. Zhao *et al*.,(2021). "Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data," in *IEEE Access*, vol. 9, pp. 5453-5465. doi: 10.1109/ACCESS.2020.3002791.

[24]. Ghorbani, R., & Ghousi, R. (2020). Comparing different resampling methods in predicting Students' performance using machine learning techniques. *IEEE Access*, *8*, 67899-67911. doi: 10.1109/ACCESS.2020.2986809.

[25]. Chui, K. T., Liu, R. W., Zhao, M., & De Pablos, P. O. (2020). Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine. *IEEE Access*, *8*, 86745-86752

[26]. Scimago. (2020). "IEEE Access". Retrieved from: https://www.scimagojr.com/journalsearch.php?q=21100374601&tip=sid&clean=0. [accessed: 15 July 2021].

[27]. Scimago. (2020). "Computers in Human Behaviour", Retrieved from: https://www.scimagojr.com/journalsearch.php?q=19419&tip=sid&clean=0.

[28]. Scimago. (2020). "Computers and Education", Retrieved from: https://www.scimagojr.com/journalsearch.php?q=17645&tip=sid&clean=0. [accessed: 15 July 2021].

[29]. Scimago. (2020). "Computers and Electrical Engineering", Retrieved from: https://www.scimagojr.com/journalsearch.php?q=18159&tip=sid&clean=0. [accessed: 18 July 2021].

[30]. Scimago.(2020). "Socio-economic Planning Sciences", Retrieved from: https://www.scimagojr.com/journalsearch.php?q=18291&tip=sid&clean=0. [accessed: 18 July 2021].

[31]. Scimago.(2020). "Materials Today: Proceedings", Retrieved from: https://www.scimagojr.com/journalsearch.php?q=21100370037&tip=sid&clean=0. [accessed: 18 July 2021].