

**2nd International Conference on  
Advanced Information Scientific  
Development (ICAISD) 2021**  
Innovating Scientific Learning for Deep Communication

Jakarta, Indonesia • 5–6 August 2021

**Editors** • Agus Junaidi, Haryani, Taufik Baidawi, Sarifah Agustiani,  
Dwi Puji Hastuti, Sopiyan Dalis and Yoseph Tajul Arifin




## Journal Development Team

Director, Publishing Development: [Bridget D'Amelio](#)

Manager, Conference Proceedings: [Emily Prendergast](#)

Editorial Assistant, Conference Proceedings: [Francesca Tangredi](#)

 For journal related inquiries, please contact: [confproc@aip.org](mailto:confproc@aip.org)

RESEARCH ARTICLE | MAY 09 2023

## Comparison of machine learning algorithm for prediction learning achievement on e-learning students

Elin Panca Saputra ✉; Indriyanti; Supriatiningsih; ... et. al



*AIP Conference Proceedings* 2714, 030005 (2023)

<https://doi.org/10.1063/5.0128538>



CrossMark

### Articles You May Be Interested In

Cervical cancer survival prediction using hybrid of SMOTE, CART and smooth support vector machine

*AIP Conference Proceedings* (April 2016)

Early detection of breast cancer in histopathology images employing convolutional neural network (CNN)

*AIP Conference Proceedings* (July 2022)

Monte Carlo's simulation method from the law of large numbers on Chebyshev's inequality

*AIP Conference Proceedings* (May 2023)



## Time to get excited.

Lock-in Amplifiers – from DC to 8.5 GHz



Find out more



# Comparison of Machine Learning Algorithm for Prediction Learning Achievement on E-learning Students

Elin Panca Saputra,<sup>1, a)</sup> Indriyanti,<sup>2, b)</sup> Supriatiningsih,<sup>3, c)</sup> Taufik Rahman,<sup>3, d)</sup>  
and Ahmad Hafidzul Kahfi<sup>4, e)</sup>

<sup>1)</sup>Informatics Technology, Universitas Bina Sarana Informatika, Jakarta, Indonesia.

<sup>2)</sup>Information Systems, Universitas Bina Sarana Informatika, Jakarta, Indonesia.

<sup>3)</sup>Computer Technology, Universitas Bina Sarana Informatika, Jakarta, Indonesia.

<sup>4)</sup>Informatics Engineering, Universitas Nusa Mandiri, Jakarta, Indonesia.

<sup>a)</sup>Corresponding author: [elin.epa@bsi.ac.id](mailto:elin.epa@bsi.ac.id).

<sup>b)</sup>Electronic mail: [indriyanti.iyt@bsi.ac.id](mailto:indriyanti.iyt@bsi.ac.id).

<sup>c)</sup>Electronic mail: [supriatiningsih.stq@bsi.ac.id](mailto:supriatiningsih.stq@bsi.ac.id).

<sup>d)</sup>Electronic mail: [taufik@bsi.ac.id](mailto:taufik@bsi.ac.id).

<sup>e)</sup>Electronic mail: [ahmad.azx@nusamandiri.ac.id](mailto:ahmad.azx@nusamandiri.ac.id).

**Abstract.** In the midst of a pandemic, online-based learning is implementing education throughout the world to keep teaching and learning running, the application of e-learning is also a mainstay to protect and prevent virus transmission. The purpose of our research is to find algorithms with high performance and classification, and to find problems in processing large data that have a high degree of accuracy by testing several algorithms contained in machine learning, namely Support Vector Machine-(SVM), neural network-(NN), Nave Bayes-(NB) and Decision Tree-(DT). The results of the algorithm that we tested, Support Vector Machine-(SVM) is the flagship that has the highest level. In addition, the advantages of Support Vector Machine-(SVM) are also found in classification modeling, because Support Vector Machine-(SVM) has a better and clearer mathematical concept than other methods. In this research, it is very important before making predictions or classifying algorithms for testing algorithms on machines that study so that algorithms that have superior performance can be found, as in this study we find Support Vector Machine-(SVM) as an algorithm that has a level performance contribution high and has a good classification. Keywords : Comparion, Machine Learning, Prediction.

## INTRODUCTION

Early predictions of student achievement, e-learning is a valuable resource for enhancing learning [1]. The Covid 19 pandemic makes e-learning or online-based learning an obligation for educational institutions around the world, because Elearning has is the obligation of all educational institutions such as schools, colleges and universities in the country and around the world due to the COVID-19 pandemic [2]. The method and process of learning by e-learning is very powerful. The power of this online learning mode can save us from this difficult time because there are many online tools available that are effective and efficient. Teachers can use a combination of audio, video, and text to reach their students [3]. This situation is challenging for education systems around the world and forces educational institutions to switch to onlinebased teaching modes [4]. Understanding e-Learning is important, so is the typology of students participating in this trend with increasing dedication [5]. One of the LMS that can be used to support online learning is Moodle, as has been implemented in several universities [6]. So thate-learning plays a major role in supporting student activities to continue to do online learning that is supported by the internet [7]. Therefore, during the pandemic, students are required to study at home using the e-learning method, the application of e-learning thus becomes an alternative technique that empowers students to be more active [8]. Because elearning can provide greater convenience to choose tutor-led or self-study courses and support them to study anytime and anywhere. However, one of the drawbacks of using e-learning is that it reduces the need for face-to-face interaction with friends [9]. So making predictions or evaluating on online learning systems is very important to determine the success of learning for students [10]. Our main research objective is to obtain the results of the level of accuracy in predicting student achievement using several algorithms in machine learning. Because study machine learning-(ML) represents an automatic model (or pattern) extraction of data. All techniques in machine learning must begin with data [11]. The method is compared with several algorithms, as well as looking to compare which algorithm has a higher level of accuracy. And this research also aims to identify and classify data from student activities that are stored following learning of e-learning. In this study, We tried to use machine-learning (ML) to get an early predictive model of student performance in completing tasks on the LMS, simply by-analyzing the resulting LMS logs files up to the time of the prediction of elearning. To get accurate results and identify the big data we

have, we have problems in making decisions so that if this big data is not handled properly, it can give incorrect results. Therefore, the approach to our research is the algorithm that is in machine learning, therefore to solve these problems it is very important we use the algorithms found in machine learning. Because machine learning can take advantage of structural patterns of data that define certain labels on observations it is known as classification because classifying data sets can solve some problems in data set classification [12]. In addition, machine learning capabilities also give machines access to large amounts of data and allow them to repeatedly perform complex mathematical calculations and improve decision support, providing a greater and faster boost [13]. Additionally machine learning is a major area of interest in the field of artificial intelligence [14]. Classifier performance appraisal is a task that can be challenging for pattern recognition, although classification knowledge is the most popular metric in the field of pattern recognition, therefore classification performance is easily evaluated unconditionally multiple metrics [15]. To get results from student achievement levels, in conducting our tests we combined several algorithms in machine learning to find a higher level of accuracy, including the Nave Bayes (NB) algorithm, the Decision Tree (DT) algorithm, the Support Vector Machine (SVM) algorithm, and the Neural Network-(NN) algorithm. In addition to finding the level of accuracy, this study is also to provide a comprehensive picture to find out the performance of each algorithm that we tested on each of the algorithm taps. From the test results we find the results of an algorithm that has a higher level of accuracy with a higher performance, namely support vector machine, because support vector machine(SVM), comparatively outperforms other algorithms in terms of increasing time and having an increase minimal error rate (17). besides Support Vector Machines-(SVM) has implications for many technical advantages which help in analyzing real-time learning analysis [16]. However, the support vector machine-(SVM) has a weakness in managing very large amounts of data, so to overcome this problem we will try to use the Smoth Support Vector Machine-(SSVM) method which is a new classification technique [17].

## **RESEARCH METHODS**

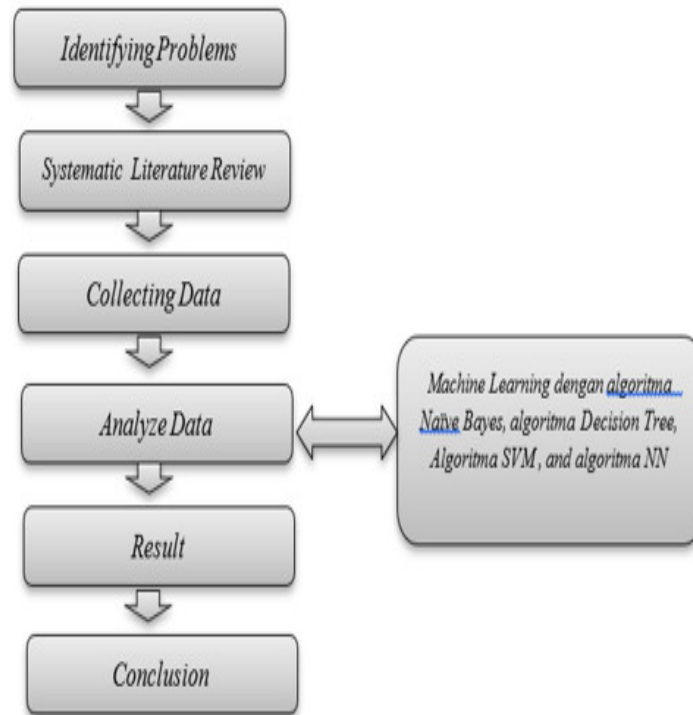
This study relies on log activity data contained in the Moodle e-learning of LMS collected by the author in the even semester of 2020 as a sample of 400 users. The approach to the methodology that we use is a combination of several methods in machine learning, including the Nave Bayes(NB) algorithm, the Decision Tree(DT) algorithm, the Support Vector Machine(SVM) algorithm, and the Neural Network-(NN) algorithm. In this study, the research data was taken through the Science Web database which was recorded from the log of student activities in conducting online teaching and learning activities. The research methodology chosen for this research is systematic. There are several stages of research that are carried out, namely identifying problems, searching for literature studies, collecting data, analyzing and writing the results of the conclusions can be seen in the image below:

### **Identifying Problems**

One of the biggest problems with the application of machine learning that we apply lies in the fact that some decision results that are generated based on various inputs are considered too complex to be described correctly. So based on this problem, machine learning can design algorithms by selecting knowledge from data based on principles and computations.

### **Systematic Literature Review (SLR)**

The next stage is the Systematic Literature Review, the literature is a secondary study carried out by mapping, identifying, evaluating in depth, combining, and collecting the results of the main study on a specific research subject [14]. Therefore SLR becomes a standard Literature Review to obtain answers by conducting a literature review based on previous studies.



**FIGURE 1.** Research methodology.

## Collecting Data

We obtained data from student teaching and learning activities from the activity log contained in the LMS-Moodle facility, we downloaded and saved the data and we grouped them one by one with a sample of 700 students who attended lectures with e-learning. And here we attach the student activity log to the Moodle LMS, which can be seen in the image below:

**TABLE I.** Log Activity From LMS 2020.

Activity Log name											
Gender	Grades report	Quiz exam	Quiz	Forum	Assignment	Chat	Message	File	Total Login	height	

## Analiza Data

Data analysis is an important part of conducting research, step by step we are doing the data analysis that we tested is the data obtained is categorized by the problem and made in the form of a systematically arranged table that makes us a research to see patterns that are related to other data.

## Result

The algorithms we use in machine learning are useful for solving problems, the results of the calculations in our research will serve as benchmarks in predicting the success rate of student achievement in e-learning, so with machine learning, we can process and analyze large and complicated data with a very short time.

## Conclusion

The conclusion of this study is that we compare several algorithms contained in machine learning, which algorithm has a level of accuracy in predicting the success rate of student learning with e-learning, because machine learning works by learning from data to improve performance on a machine or system.

## TESTING THE CALCULATION RESULTS OF SEVERAL ALGORITHMS

Here we describe the results of the calculations of some of the algorithms that we tested

### Processing results calculated by Support Vector machine

From the management of Support Vector Machine calculations from confusion matrix can be visited in table 3. With a True Positive- (TP) score of 256 which will be classified as 1 afterwards with predictions calculated by Neural Network- (NN), and the total value is False-Negative 65 data will be predicted with 1 but 2, after that the number of True-Negative results with the number of 6 data which is used for 2 is in accordance with the prediction, and False-Positive with a total of 75 data which is predicted to be 2 but becomes 1. then The value of the accuracy that has been calculated by the Support Vector Machine is 82.25%, while the AUC has a value of 85.80 and for details can be seen in Table 2.

TABLE II. Result of Calculation and data Examiner (2021)

Classification	Predicted-Class	
	Class for 1	class for 2
class for 1	256	65
class for 2	6	73

### Processing results calculated by Neural Network

From the management of Neural Network calculations from confusion-matrix can be visited in table 3. With a True Positive- (TP) score of 229 which will be classified as 1 afterwards with predictions calculated by Neural Network- (NN), and the total value is False-Negative 51 data will be predicted with 1 but 2, after that the number of True-Negative results with the number of 33 data which is used for 2 is in accordance with the prediction, and False-Positive with a total of 87 data which is predicted to be 2 but becomes 1. then The value of the accuracy that has been calculated by the Neural-Network is 79.00%, while the AUC has a value of 81.50 and for details can be seen in Table 3.

TABLE III. Result of Calculation and Data Examiner (2021)

Classification	Predicted-Class	
	Class for 1	class for 2
class for 1	229	51
class for 2	33	87

### Processing results calculated by Naïve Bayes-(NB)

From the management of Naïve Bayes calculations from confusion-matrix can be visited in table 3. With a True Positive- (TP) score of 251 which will be classified as 1 afterwards with predictions calculated by Neural Network- (NN), and the total value is False-Negative 80 data will be predicted with 1 but 2, after that the number of True-Negative results with the number of 11 data which is used for 2 is in accordance with the prediction, and False-Positive with a total of 58 data which is predicted to be 2 but becomes 1. then The value of the accuracy that has been calculated by the Naïve Bayes is 77.25%, while the AUC has a value of 75.70 and for details can be seen in table 4:



TABLE IV. Result of Calculation and Data Examiner (2021)

Classification	Predicted-Class	
	Class for 1	class for 2
class for 1	251	80
class for 2	11	58

### Processing result calculated by Decision Tree-(DT)

From the management of Decision Tree calculations from confusion-matrix can be visited in table 3. With a True Positive- (TP) score of 255 which will be classified as 1 afterwards with predictions calculated by Decision Tree, and the total value is False-Negative 70 data will be predicted with 1 but 2, after that the number of True-Negative results with the number of 7 data which is used for 2 is in accordance with the prediction, and False-Positive with a total of 68 data which is predicted to be 2 but becomes 1. then The value of the accuracy that has been calculated by the Decision Tree is 80.75%, while the AUC has a value of 75.20 and for details can be seen in table 5:

TABLE V. Result of Calculation and Data Examiner (2021)

Classification	Predicted-Class	
	Class for 1	class for 2
class for 1	255	70
class for 2	7	68

## EVALUATION RESULTS COMPARISON OF SEVERAL ALGORITHMS THAT HAVE BEEN TESTED WITH CALCULATION

For details, the results of the testing of several algorithms that we did to get 4 algorithms were accurate, namely Neural Network Support, Vector Machine, Naive Bayes and Decision Tree, for details, please visit table 6.

TABLE VI. Comparison Result (2020)

Algorithm	Accuracy	AUC	Classification Result
Support Vector Machine-(SVM)	82.25%	85.80%	good classification
Neural Network-(NN)	79.00%	81.50%	fair classification
Naive Bayes-(NB)	77.25%	75.70%	fair classification
Decision Tre-(DT)	80.75%	75.20%	good classification

## CONCLUSION

If you look at table 6, the results of the comparisons of several algorithms that have been tested in performing calculations are the results with the calculation of Support Vector Machine-(SVM) with a score of 82.25% and AUC of 85.80%, Neural Network-(NN) with a value of 79.00% and AUC 81.50%, Naive Bayes- (NB) with a score of 77.25% and AUC of 75.20%, and the Decision Tre-(DT) algorithm with a score of 80.75%, and AUC 75.20%. It can be concluded that from several algorithms that have been tested for calculations, the Support Vector Machine-(SVM) algorithm has a higher level of accuracy than the Neural Network-(NN), Naive Bayes- (NB) and Decision Tre-(DT) algorithms. So it is clear that the advantages of Machine Learning on Support Vector Machine-(SVM) are superior to other algorithms, because Support Vector Machine-(SVM) is part of a supervised learning models that has a relationship with data analysis and pattern recognition. Therefore the Support Vector Machine method is to take a set of input data and then estimate for each given input from the two possible classes to make output. Therefore the SVM algorithm is to take a set of input data and then estimate for each given input from the two possible classes to make output. But there are problems with us if in the future we still use the SVM algorithm using large-scale data. So to solve this problem for our next research, namely with large data, we will develop with Smoth Support Vector Machine and Reduced Support Vector Machine.



## REFERENCES

1. M. Riestra-González, M. del P. Paule-Ruíz, and F. Ortin, "Massive LMS log data analysis for the early prediction of course-agnostic student performance," *Comput. Educ.*, vol. 163, p. 104108, Apr. 2021, doi: 10.1016/j.compedu.2020.104108.
2. R. Radha, K. Mahalakshmi, V. S. Kumar, and A. R. Saravanakumar, "E-Learning during Lockdown of Covid-19 Pandemic: A Global Perspective," *Int. J. Control Autom.*, vol. 13, no. 4, pp. 1088–1099, 2020.
3. S. Dhawan, "Online Learning: A Panacea in the Time of COVID-19 Crisis," *J. Educ. Technol. Syst.*, vol. 49, no. 1, pp. 5–22, 2020, doi: 10.1177/0047239520934018.
4. O. M. Gushchina and A. V. Ochepovsky, "Data mining of students' behavior in E-learning system," *J. Phys. Conf. Ser.*, vol. 1553, no. 1, 2020, doi: 10.1088/1742-6596/1553/1/012027.
5. A. Topîrceanu and G. Grossec, "Decision tree learning used for the classification of student archetypes in online courses," *Procedia Comput. Sci.*, vol. 112, pp. 51–60, 2017, doi: 10.1016/j.procs.2017.08.021.
6. K. Pisutova, "Innovation of elearning at comenius university," *ICETA 2019 - 17th IEEE Int. Conf. Emerg. eLearning Technol. Appl. Proc.*, pp. 644–648, 2019, doi: 10.1109/ICETA48886.2019.9039958.
7. T. Favale, F. Soro, M. Trevisan, I. Drago, and M. Mellia, "Campus traffic and e-Learning during COVID-19 pandemic," *Comput. Networks*, vol. 176, no. May, 2020, doi: 10.1016/j.comnet.2020.107290.
8. F. Rahmawati, "E-Learning Implementation: Its Opportunities and Drawbacks Perceived by EFL Students," *J. Foreign Language Teach. Learn.*, vol. 1, no. 1, 2016, doi: 10.18196/ftl.111.
9. J. @ E. Luaran, N. N. Samsuri, F. A. Nadzri, and K. B. M. Rom, "A Study on the Student's Perspective on the Effectiveness of Using e-learning," *Procedia - Soc. Behav. Sci.*, vol. 123, pp. 139–144, 2014, doi: 10.1016/j.sbspro.2014.01.1407.
10. E. P. Saputra, Supriatiningsih, Indriyanti, and Sugiono, "Prediction of Evaluation Result of E-learning Success Based on Student Activity Logs with Selection of Neural Network Attributes Base on PSO," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012074.
11. B. Kumar, O. P. Vyas, and R. Vyas, "A comprehensive review on the variants of support vector machines," *Mod. Phys. Lett. B*, vol. 33, no. 25, pp. 1–11, 2019, doi: 10.1142/S0217984919503032.
12. J. Faouzi and H. Janati, "Pyts: A python package for time series classification," *J. Mach. Learn. Res.*, vol. 21, pp. 1–6, 2020.
13. M. Attaran and P. Deb, "Machine Learning: The New 'Big Thing' for Competitive Advantage," *Int. J. Knowl. Eng. Data Min.*, vol. 5, no. 1, p. 1, 2018, doi: 10.1504/ijkedm.2018.10015621.
14. Y. Xu, Y. Zhou, P. Sekula, and L. Ding, "Machine learning in construction: From shallow to deep learning," *Dev. Built Environ.*, vol. 6, no. February, p. 100045, 2021, doi: 10.1016/j.dibe.2021.100045.
15. O. Aydemir, "A New Performance Evaluation Metric for Classifiers: Polygon Area Metric," *J. Classif.*, 2020, doi: 10.1007/s00357-020-09362-5.
16. A. Aktas and M. Kabak, "A Hesitant Fuzzy Linguistic Group Decision Making Model for Energy Storage Unit Selection," vol. 1197 AISC, 2021.
17. M. Y. Darsyah, I. J. Suprayitno, F. Fuzi, B. W. Otok, and B. S. S. Ulama, "Smooth Support Vector Machine (SSVM) for classification of human development index," *J. Phys. Conf. Ser.*, vol. 1217, no. 1, 2019, doi: 10.1088/1742-6596/1217/1/012114.