Machine Learning Algorithms for Early Predicting Dropout Student Online Learning

2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED) | 979-8-3503-7012-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCED60214.2023.10425359

Meta Amalya Dewi
Information Systems Department
BINUS Online Learning
Bina Nusantara University
Jakarta, Indonesia 11480
meta.dewi@binus.edu

Felix Indra Kurniadi
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
felix.kurniadi@binus.edu

Dina Fitria Murad
Information Systems Department
BINUS Online Learning
Bina Nusantara University
Jakarta, Indonesia 11480
dmurad@binus.edu

Sucianna Ghadati Rabiha
Information Systems Department,
BINUS Online Learning
Bina Nusantara University
Jakarta, Indonesia 11480
Sucianna.rabiha@binus.ac.id

Awanis Romli
Faculty of Computing,
Universiti Malaysia Pahang,
Pahang, Malaysia
awanis@ump.edu.my

***Abstract*— Online learning is different from offline learning in the classroom with supervision from the lecturer. Online learning using the Learning Management System (LMS) media requires high awareness from students because their learning activities are not supervised, they are free to study wherever and whenever, so they need to manage and control their own study time without the help of lecturers or administrators. This is one of the causes of the high dropout rate among online learning students, so it is very important for lecturers and administrators to support students in a timely manner to avoid the risk of dropping out. This study uses access log data recorded in the LMS and student statistical information and calculated data and aims to present a suitable predictive algorithm for dropout early prediction systems for online learning students using machine learning. Of the 4 algorithms used, the highest recall value is in Naïve Bayes (1), the highest precision is in Logistic Regression with Lasso (1), while the highest accuracy value (0.99) and F1score (0.97) are obtained from the Support Vector Machine which has value equal to Logistic Regression with Lasso. In general, the early dropout prediction model will allow lecturers and administrators to focus on students who have the potential to dropout and take quick action to improve their learning performance so as to reduce the number of student dropouts.**

Keywords*— *Predictions, student, dropout, online learning, machine learning.

I. INTRODUCTION

Online learning is the best choice for students who are busy working to learn new knowledge and earn a degree without any time or space constraints. And when the SARS-Cov-2 virus spread rapidly to become a pandemic, all universities changed offline or face-to-face learning methods to online to avoid transmission of the virus between students and teachers [1]. Given the time management of students and unsupervised educational activities in online learning, they are required to have the responsibility to complete their studies independently, in fact there are many cases of students dropping out [2], of course this is a challenge for universities providing online learning. Active intervention from lecturers or managers is needed for students who have the potential to dropout.

For this reason, universities providing online learning must improve the analytic system for learning so that students who have the potential to drop out of school can be found as early as possible and then provide appropriate interventions obtained from the results of the analysis of student behaviour recorded in the LMS [3]. This prediction helps lecturers or administrators take appropriate actions to improve learning outcomes, so as to increase student graduation rates, help universities make appropriate learning strategies, and provide critical feedback to students and lecturers [4].

The prediction algorithm is used to produce predictions with a high degree of accuracy in identifying students who have low achievements and have the potential to dropout. It takes student datasets from LMS to be analysed and some machine learning models are used for predictions. Machine learning offers advantages over traditional statistical analysis methods. Predictive models are developed using machine learning methods and available data are used for plot patterns in the decision-making process [5]. Using mathematical and statistical procedures, machine learning is a software modelling technique for self-learning systems that draws conclusions from data or experiences [6].

Online learning is a way of learning by applying the using the internet to obtain learning resources, discuss with lecturers and other students, and receive assistance during the learning process [7]. The focus of online learning is student attitudes, perceptions, assessment evaluations, satisfaction and learning performance [8]. Today, online learning does not talk about the wishes of many parties with their own goals, but it has become a necessity [9].

According to [10], there are factors that affect students' ability to participate in digital learning, including difficult to reach material, lack of technology, internet availability as well as the lack of student interaction both with teachers and other students are a big influence on the implementation of online learning.

Universities that offer online learning must contend with rising dropout rates due to low student retention [2]. Dropout causes a person to become deprived of education which has an impact on a lack of social and economic welfare [11]. In addition, the inability of a nation to produce as much as it