

An Investigation into Student Performance Prediction using Regularized Logistic Regression

Felix Indra Kurniadi
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia, 11530
felix.indra@binus.ac.id

Meta Amalya Dewi
Information Systems Department
BINUS Online Learning
Bina Nusantara University
Jakarta, Indonesia, 11480
meta.dewi@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, 26600
awanis@ump.edu.my

Abstract—The problem of university dropout poses a significant challenge to education systems worldwide, affecting administrators, teachers, and students. Early identification and intervention strategies are crucial for addressing this issue. In addition, advances in machine learning have paved the way for more accurate predictions of student performance. This paper investigates the use of regularization techniques, specifically Lasso (L1) and Ridge (L2) regularization, within logistic regression models to improve the accuracy of performance prediction. This research's dataset was obtained from the Binus Online Learning platform at Bina Nusantara University, with a focus on the Information System study program between 2020 and 2021. The results reveal that logistic regression with regularization achieves a high level of accuracy, recall, and precision in predicting student performance. The findings contribute to the development of an early warning system to identify at-risk students, enabling timely intervention and support.

Keywords—student performance score, logistic regression, regularization, Lasso regression, Ridge Regression

I. INTRODUCTION

The issue of university dropout poses a significant challenge to education systems around the globe, affecting academic and financial administrators, educators, and students. The consequences of students dropping out of college extend beyond individual circumstances to include economic and societal repercussions. In addition, the failure and frustration related to dropping out contribute to affected students' negative self-perception [1].

Education stakeholders must take proactive measures to address the urgent problem of university dropout and its far-reaching ramifications. By recognizing the obstacles encountered by students at risk of dropping out and implementing early intervention strategies, the education system can attempt to mitigate the negative effects of this phenomenon. Early

identification of students' academic performance can aid educators in identifying those who may require supplementary courses, additional assignments, or additional support [2].

Machine Learning is a rapidly developing field at present, and its advancements are also applicable to the field of education, particularly in determining student performance. This can be observed through the implementation of early detection of student performance using machine learning approaches [3]–[6].

Several researchers try to identify Students At-Risk and Minimizing Failure in University using Machine Learning. Segura et al. conducted a study on the negative effects of university attrition rates. Using data from a large face-to-face university in Europe, the researchers focused on identifying potential dropouts early on. The purpose of the study was to predict the likelihood of students dropping out at the beginning of the course or by the end of the first semester by analyzing five main program areas. Utilized techniques included feature selection and machine learning models including Support Vector Machines, Decision Trees, Artificial Neural Networks, and Logical Regression. Findings revealed that attrition prediction is not solely dependent on enrollment variables, but improves when first-semester results are considered. Other factors, such as a student's level of interest in their selected course, also played a role, despite the significance of academic performance. The success of prediction techniques varied across program areas, with machine learning models achieving the greatest results, while a simple logistic regression model served as an adequate baseline [1].

Kanchana et al. emphasized the significance of quality education as an objective for sustainable development, as well as the role of technology and affordable internet access in expanding educational opportunities. They emphasized the