Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015-2021)

https://doi.org/10.3991/ijet.v17i05.27685

Muhammad Haziq bin Roslan, Chwen Jen Chen
Universiti Malaysia Sarawak, Kota Samarahan, Malaysia
cjchen@unimas.my

Abstract—This systematic literature review aims to identify the recent research trend, most studied factors, and methods used to predict student academic performance from 2015 to 2021. The PRISMA framework guides the study. The study reviews 58 out of 219 research articles from Lens and Scopus databases. The findings indicate that the research focus of current studies revolves around identifying factors influencing student performance, data mining (DM) algorithms performance, and DM related to e-Learning systems. It also reveals that student academic records and demographics are primary aspects that affect student performance. The most used DM approach is classification and the Decision Tree classifier is the most employed DM algorithm.

Keywords—Educational Data Mining (EDM), data mining (DM) techniques, prediction studies, student academic performance, systematic literature review

1 Introduction

Student academic performance is a significant aspect in determining educational success at all levels [1, 2, 3]. Academic performance is crucial for students to continue their studies and secure their future [4, 5]. Several studies uncover factors that can predict their performance. Earlier studies such as McKenzie and Schweitzer [6], Andujar et al. [7], and Taniguchi et al. [8] employed classical statistical methods to determine these factors. Generally, regression analysis, discriminant analysis, and cluster analysis are examples of classical statistics approaches used in this area [9]. Artificial intelligence methods such as Backpropagation, Support Vector Regression, Gradient Boosting Classifier [10], Bayesian Classifier, Artificial Neural Network, and Decision Tree [11] are later employed. The latter involves a mix of advanced statistical methods and artificial intelligence heuristics and has contributed to the growth of educational data mining (EDM), which adds to our understanding of how to predict student academic performance [12, 13, 14, 15, 16]. EDM research studies [17] applied data mining (DM) techniques to data obtained from diverse educational systems to improve the quality of education [18]. Such mining is significant as it enables educators to take necessary interventions to achieve optimal student performance [19].

To date, scholars are debating how to anticipate students’ academic performance, focusing on the types of variables that influence such a prediction [20, 21]. Based on a