Integration of Educational Data Mining Models to a Web-Based Support System for Predicting High School Student Performance

Authors : Sokkhey Phauk, Takeo Okazaki

Abstract : The challenging task in educational institutions is to maximize the high performance of students and minimize the failure rate of poor-performing students. An effective method to leverage this task is to know student learning patterns with highly influencing factors and get an early prediction of student learning outcomes at the timely stage for setting up policies for improvement. Educational data mining (EDM) is an emerging disciplinary field of data mining, statistics, and machine learning concerned with extracting useful knowledge and information for the sake of improvement and development in the education environment. The study is of this work is to propose techniques in EDM and integrate it into a web-based system for predicting poor-performing students. A comparative study of prediction models is conducted. Subsequently, high performing models are developed to get higher performance. The hybrid random forest (Hybrid RF) produces the most successful classification. For the context of intervention and improving the learning outcomes, a feature selection method MICHI, which is the combination of mutual information (MI) and chi-square (CHI) algorithms based on the ranked feature scores, is introduced to select a dominant feature set that improves the performance of prediction and uses the obtained dominant set as information for intervention. By using the proposed techniques of EDM, an academic performance prediction system (APPS) is subsequently developed for educational stockholders to get an early prediction of student learning outcomes for timely intervention. Experimental outcomes and evaluation surveys report the effectiveness and usefulness of the developed system. The system is used to help educational stakeholders and related individuals for intervening and improving student performance.

Keywords : academic performance prediction system, educational data mining, dominant factors, feature selection method, prediction model, student performance