Using Predictive Analytics to Identify First-Year Engineering Students at Risk of Failing

Authors : Beng Yew Low, Cher Liang Cha, Cheng Yong Teoh

Abstract : Due to a lack of continual assessment or grade related data, identifying first-year engineering students in a polytechnic education at risk of failing is challenging. Our experience over the years tells us that there is no strong correlation between having good entry grades in Mathematics and the Sciences and excelling in hardcore engineering subjects. Hence, identifying students at risk of failure cannot be on the basis of entry grades in Mathematics and the Sciences alone. These factors compound the difficulty of early identification and intervention. This paper describes the development of a predictive analytics model in the early detection of students at risk of failing and evaluates its effectiveness. Data from continual assessments conducted in term one, supplemented by data of student psychological profiles such as interests and study habits, were used. Three classification techniques, namely Logistic Regression, K Nearest Neighbour, and Random Forest, were used in our predictive model. Based on our findings, Random Forest was determined to be the strongest predictor with an Area Under the Curve (AUC) value of 0.994. Correspondingly, the Accuracy, Precision, Recall, and F-Score were also highest among these three classifiers. Using this Random Forest Classification technique, students at risk of failure could be identified at the end of term one. They could then be assigned to a Learning Support Programme at the beginning of term two. This paper gathers the results of our findings. It also proposes further improvements that can be made to the model.

Keywords : continual assessment, predictive analytics, random forest, student psychological profile