

## **Recommendations for Students in Higher Education: A Machine Learning Approach**

K.T.S. Kasthuriarachchi<sup>1</sup>, S.R.Liyanage<sup>1</sup>

Educational Data Mining is a rising discipline in Data Mining setting which concentrated on creating systems for investigating one of a kind data that starts from educational settings, and utilizing those procedures to better comprehend students and the settings which they learn in. There were numerous potential circumstances for applying data mining in education, such as; predicting the performance of students in education domain, advancement of student models, making methodologies for instructive help, settling on decisions to growing better learning systems, upgrading the execution of students and lessening the dropout rate of students and so on. There were sure examinations directed in dissecting students' data to foresee the execution in light of data mining approaches utilizing machine learning algorithms. However, a few of them were guiding the students using the recommendations of educators to success in their academic lives. The key objective of this research is to provide educators' recommendations to students in higher education through data analysis using machine learning algorithms. In this experiment, the data about more than 3000 students with eight attributes; age, gender, A/L Stream, A/L English Grade, does the student has repeat modules, GPA of Semester1, GPA of Semester 2 and Pass status of year 1 were included into the research sample who registered and were following their first academic year of an Information Technology degree in an institute. Three classification type machine learning algorithms were used to build the predictive model. They were Naïve Bayes algorithm, Decision Tree algorithm and Support Vector Machine algorithm. The accuracy of the models built by each algorithm have been tested against each other to identify the best model and extracted the most influencing/ important attributes in the model to predict the final grade (pass/fail) in the end of first year of the students. Accordingly, the accuracy measures of Naïve Bayes, Decision tree and Support Vector Machine were recorded as 74.67%, 74.01% and 74.01% respectively and it was clear that all three algorithms were holding almost same accuracy level. However, the model generated by Naïve Bayes algorithm has been selected since it was outperformed the rest. Then rank features by importance method was used as the feature selection method to identify the most influencing factors of the predictive model. As the result of it, past repeat modules, GPA of Semester1, GPA of Semester 2 were extracted as the most influencing attributes. Furthermore, these attributes were tested using correlation analysis to measure the significance of the relationship with the target attribute. According to this study, the educators will be able to recommend the students to score good marks for assessments of the subjects to obtain a better GPA to semester 1 and semester 2 without failing the modules to successfully complete the first year of the degree course which make more beneficial for educators as well as students to be success.

**Keywords:** Educational Data Mining, Machine Learning, Accuracy, Feature selection, Grade Point Average

---

<sup>1</sup>University of Kelaniya, Sri Lanka