

A MACHINE LEARNING APPROACH FOR UNIVERSITY STUDENT AND COURSE PROFILING WITH RESULT PREDICTION: A CASE STUDY

T.R.N. Perera^{*} and E.O. Dayaratna

Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.

**rumalip95@gmail.com*

A wide variety of courses is offered at state universities in Sri Lanka, and students have the freedom of selecting subjects and courses they prefer under the selected study programme. In the first two years in the Faculty of Science, University of Peradeniya, Sri Lanka, students have to register for many compulsory courses; but in the next two years, they have a freedom of selecting optional courses. Even though students have three major subjects in the first year, they are given the opportunity to register for courses from many other disciplines to fulfil the credit requirement within the academic year. A proper mechanism for a student to select the best elective courses according to his or her capabilities is imperative to elevate the student's academic performance. Thus, the aim of this research is to find an effective way of identifying best elective courses suitable for a student to improve the academic performance. The proposed method consists of two approaches to solve this problem. In the first approach, using the results of past students for various courses, the grade a new student might obtain for a course is predicted. A dataset of 208,985 records is considered for this study from the year 2005 up to 2012 which was received from the Faculty of Science, University of Peradeniya. Pre-processing this dataset was carried out by removing non-processable values and representing values in one standard format and adding new attributes such as the Grade Point value corresponding to each grade for different analyses. Machine learning techniques, such as the K-nearest neighbour method, neural networks, decision trees and the random forest method were used for the prediction of results. Neural networks is identified to be the best machine learning technique which gives the highest and the most reliable accuracy of 99.69% when using the accuracy score function for results prediction. Using this method, students can predict their results before attempting any course. The second approach of this research was to identify various course profiles/groups of similar courses offered at the Faculty. Using the K means algorithm, keywords from course descriptions and pre-requisite courses, the courses were categorized into ten clusters. With the help of these course profiles/groups as a base, student profiles/groups were identified. If a student performs well for a particular course, its course group can be identified, and the student can select any other course from this group as elective courses. These selected elective courses are predicted to give similar results as the former course. Thus, the proposed method in this study is recommended to identify and select the best elective courses according to the student's capabilities.

Keywords: Course clusters, Machine learning approach, Results Prediction, Student profiles