Title: Adaptive E-learning
PHYS1160 Case Study

Author: Jessica Ai
Supervisor: Dr. Nadine Marcus, Dr. Michael Bain

Aim of Research

The premise of this study is to add to the present literature within Educational Data Mining by analysing student behaviour pertaining to a fully online university course, titled Introduction to Astronomy (PHYS1160). This course is offered to first year physics students and also as a general elective available for older students from other faculties. The semester under examination in this paper was the first session run since the redesign of the course which now makes use of adaptive e-learning platform, Smartsparrow along with the popular course management system, Moodle. A matched dataset across these two platforms was obtained for this study. The analysis is approached with the goal of using machine learning and statistical techniques to generally explore model student behaviour and more specifically examine the observable aspects that can strongly influence student performance. The motivations are to provide insights to help identify students at risk of performing poorly with the hope that intervention strategies can be framed around them. The study found that the number of interactions with the discussion forum.

Hypothesis

The dataset consisted of Moodle navigational click data, student grades and Smartsparrow activity scores. The hypothesis was that the number of interactions with the two platforms could be used as a proxy for student engagement, which in relevant literature has been linked to student performance. It was found that the analysis was initially too granular, and this hypothesis was refined to predicting student bands based on data aggregated at a weekly level.

Methodology

Cluster centroids were defined as the weekly average level of discussion forum interactions for each mark band. For each week, students were then clustered according to the number of standard deviations away from each of the centroids. Then, the cumulative average for each week was calculated to obtain the eventual mark band predictions for each week. The baseline measure used was the initial distribution of the mark band distributions.

Results

The first figure to the right shows the average level of the number of interactions with the online course platforms across the 16 weeks of the semester, for each mark band. It is clear that HD students consistently interact more with the platform, and reading the chart from top to bottom follows the order of the mark bands. It was found that student's interactions with the forums consisted showed strong correlation, across all weeks, with the student's final grade. As a result, the methodology outlined above was used to predict the student's final grade and the accuracy by week is shown in the chart below, where you can see that the prediction accuracy peaks by week 8. Furthermore, a confusion matrix is presented in the bottom right along with the % improvement over the baseline measure for the week 8 classification results. It is unfortunate that the prediction fails to correctly predict students who fail and it is recommended that any intervention strategies be aimed towards those who are predicted to pass instead.

Conclusion

It is of course difficult to generalise what we learn from one course to another and results on the predicting academic performance in the existing literature vary from one case study to another. The research presented here demonstrates that there is a strong relationship between student engagement and academic performance, and this is in line with the existing literature. However, there were several limitations in the dataset, Moodle's web logs did not provide sufficient information, particularly in that it did not have the length of time a webpage was active, and granularity in the Smartsparrow dataset was lost after matching was performed. In general, if online learning is to succeed where traditional classrooms have failed, in providing different pathways to accommodate different learning styles, then more meaningful data needs to be recorded such that it can inform future work.