Predicting student success based on interactions with Virtual Learning Environment.

Vivek Doijode, Neha Singh
MS in Business Analytics, Oklahoma State University
Introduction

• Online learning can be called the millennial sister of classroom learning; tech savvy, always connected, and flexible.

• These features offer a convenient alternative to students with constraints and working professionals to learn on demand.

• According to National Center for Education Statistics, over 5 million students are currently enrolled in distance education courses.

• The growing trend and popularity of MOOCs (Massive Open Online Courses) and distance learning makes it an interesting area of research.
Introduction...

• We plan to work on OULA (Open University Learning Analytics) dataset.
• Learning analytics provides many insights on the learning pattern of students and on module assessments.
• These insights may be researched to enhance participants’ learning experience.
• In this paper, we predict students’ success in an online course using regression, clustering and classification methods.
Introduction...

• We have a mix of categorical and numeric inputs present in the OULA datasets that are in csv file formats and contain information for more than 30,000 students pertaining to 7 distance learning courses, student demographics, course assessments and student interaction with virtual learning environment.

• We have merged tables together using unique identifiers. We will first explore the merged data using SAS to generate insights and then build appropriate predictive models.
Data Preparation and Analysis:

- Data is obtained by joining 7 different tables (Fig 1). The Student Info table contains demographic details of students, Student Registration contains information on when the students registered/unregistered for the courses, StudentVLE and VLE tables contain virtual learning environment information, Student Assessment, Assessment tables contain information on assessments.
Data Preparation and Analysis:

- Information from VLE tables is summarized to get the total sum clicks for various types of activities the student undertakes for a course module. Each student undergoes several assessments over the duration of course. Assessments are weighted and students may opt to drop out of courses by withdrawing.

Fig. 2 Percentage of Pass/Fail/Withdrawn by region
Fig. 3 Module wise percentages of Pass/Fail/Withdrawn

Fig 4. Frequency of final result for assessments

Fig 5. Sum Clicks by each final result
Decision Trees

• Using decision tree, we would be able to explain the most important variables of our analysis by observing the top segment of the decision tree and analyzing the variable importance matrix.

• Fig 7. represents an interactive decision tree model which uses the classifying variables based on their (logworth) value.
Decision Trees..

- Misclassification rate of the train and validation data is close and is less as 0.10, so that is a fair estimate of a good model but if we look at the data split, this split results into homogenous groups of withdrawn and pass students which at this point is not desirable.

- Hence we ignore this variables and instead do our first split on the second most important variable i.e latest_date_of_interaction.
Conclusions

• The most important variables are latest_date_of_interaction that is measured as number of days relative to the start of module presentation, followed by Score, Code_module.

• Students who scored greater than 57 in initial assessments had a high pass percentage of 90.01 as compared to the other group who scored less than 57.

• Of the students who scored more than 73.5 and had total clicks greater than 1352 had the greatest pass percentage of 96.14 whereas only 3.4% ended up failing.

• Of the students who scored less than 73.5 and with clicks less than 1352, 8.16% ended up failing the course.
Future Work

• The scope of this project will be extended to do back test of the model and implement the successful validation results to identify students’ at risk, applicable to online course websites such as Coursera, Udacity.