Introduction

Since 2020, there has been a major educational shift out of the classroom and into virtual learning environments (VLE). In the past decade alone, the proportion of students enrolling in online education nearly quadrupled (Hamilton 2024). Although online education provides students with more flexible access to the curriculum and educational administrators, educators find it difficult to gauge how well the material is transferring to their students. The separation of student to educator removes much of the nonverbal communication from the classroom, only leaving hard data behind.

Thanks to Dr. Kuzilek and their colleagues through The Open University (2015), online academic program data is being made available to understand driving factors for successful students in VLEs. The data provided could simultaneously be used to train classification models that could help administrators predict whether students are going to meet the Student Learning Outcomes. Although the models trained out of this data set may be out of date, they may provide the framework to build more complex and robust tools for other programs or times.

Data Wrangling

The Open University (2015) data set depicts the outcomes of seven online courses (modules) from The Open University. Seven separate comma separated value (csv) files were provided as the complete data set. In total, there were 32953 students that worked on 206 different assessments require various levels of interaction with the VLE. The data set provides the demographics of the students, their individual interactions with each activity as part of the assessments, and the classification of how they ended the course: Distinction, Pass, Fail, or Withdrawn. Distinction is a higher classification than passing. No final scores were provided in any of the csvs.

The assessments csv characterized the assessments the students were being scored on. Characterizations of the assessments included which module they belonged to (‘code\_module’), assessment identifiers (‘id\_assessment’), the due date (‘date’), and the impact of the assessment on the overall grade (‘weight’). The only feature that was used from the assessments csv was the final submission date. There were a few assessments that were missing submission dates that were imputed with the maximum submission date. The maximum submission date represents the last day of the course since all dates are formatted as number of days since the start of the module. By imputing the last day of the course, the students are assumed to not be able to submit past the end of the module. All other features were either used for merging the DataFrames or dropped all together.

The courses csv provided the module identifiers as well as the length of the course. All features from this DataFrame were dropped after merging because the focus of this model is features directly related to or within the students’ control.

The studentAssessment csv showed how well each student scored on their respective assessments (‘score’) and when they turned in each assessment (‘date\_submitted’). Because some students withdrew from the courses, not every assessment was completed by every student. Administrative features like whether the assessment was transferred, or the unique identification number (ID) were dropped prior to model construction. The DataFrame was missing about 0.09% of the observations' scores. With such a small proportion missing of a large dataset, the observations were dropped.

The studentInfo csv showed the demographic information of each student. Students were only distinguished by a unique ID (‘id\_student’). Even the students’ ages were classified within age bands (‘age\_band’) to further protect their identities. Administrative features like ID were dropped prior to model construction and analysis. The main features that were provided within this DataFrame included the final result (‘final\_result’) of the student with the module, the highest level of education attained (‘highest\_education’), and the number of additional credits being studied by the student (‘studied\_credits’), among other demographic information. The Index of Multiple Depravation (IMD) band establishes the financial bracket of the students. Missing values in the IMD band feature (‘imd\_band’) were imputed with ’20-30%’ which accounted for 12% of the feature’s data initially.

The studentRegistration csv provided the date each student registered for each module; and if they withdrew, what date they unregistered. Only the registration date (‘date\_registration’) was retained of all the features to see if there was a relationship between when a student registered with how well they did in the course. The dates were all relative to the start of the module. Negative dates indicated that the student registered prior to the start of the course.

The studentVle csv was used for the number of days the students interacted (‘date’) with the VLE and how many clicks they performed on each of these days (‘sum\_click’). All other features were administrative in nature that were only useful for merging. The number of clicks were of special interest as a quantification of student engagement with the VLE.

Finally, the vle csv was used to characterize the courses with the various VLE activities. Only the types of activities (‘activity\_type’) were retained after merging to see if any part of the VLE was associated with how students finished the course.

All duplicates were dropped and the only feature that was simply renamed was ‘date\_submitted’ to ‘assessment\_duration’. The date provided by this feature can be used as a duration because all dates are relative to the start of the module.

Aggregations

While 'date' is a duration length relative to the start of the module, it only served as a marker for each day of interaction with each activity the students interacted with per assessment so it was aggregated as a count of the number of days interacting with the material to quantify the interaction with each activity type (‘days\_active’). Similarly, 'sum\_click' was aggregated to the total number of clicks per activity (‘total\_clicks’).

The number of clicks were further aggregated as the total number of clicks per assessment (‘clicks’) and the average number of clicks for each assessment (‘mean\_clicks'). The number of active days were also aggregated per assessment into the sum total (‘total\_active’) and average (‘mean\_active’). As two distinct quantifications of student interaction with the material.

The scores were also averaged (‘mean\_score’) for each assessment to represent the consistency and level that students operated on. The ‘assessment\_duration’ was also averaged (‘mean\_assessment\_length’) and maxed (‘max\_assessment\_length') to represent how long students had with the material.

Data Exploration

The two main goals outlined for this model is to identify the key drivers of how students ended the course and create a classification model that enables educators to predict student trajectories. In relation to these goals, four questions were used to initialize understanding student VLEs.

1. How does the average score per assessment differ between the final results?
2. How does the material interaction for each assessment differ between the final results?
3. Is there a difference between activity types that determine the final results?
4. Does education interact with students' final results?

How does the average score per assessment differ between the final results?

One of the most important predictors that would dictate a student's final result is how well they scored on average with their assessments. A student that scores higher on average per assessment should place higher with a pass or even a distinction final result since they would have met the minimum passing grade per assessment at least.

A group of green and white graphs

Description automatically generated

Fig 1. Distribution of the average score for each assessment submitted by all students distinguished by how they ended the courses.

Lower ranked student results are more normally distributed than distinguished or passing students (Fig 1). Despite having a difference in a couple of points there is a significant difference (f-stat = 8332.65, p-value = 0.0) between the distribution means of students who failed (mean score = 65.14) and withdrew (mean score = 63.74). Even with Bonferroni correction, all four classifications of students were tested to not be like one another (f-stat = 8332.65, p-value = 0.0).

From the initial testing, the average score for each assessment appeared to be a driving factor in classification. Students who, on average, scored higher were more likely to pass or be distinguished than those who scored lower.

How does the material interaction for each assessment differ between the final results?

Increasing the amount spent with the material may indicate a student that is taking special care with the assessment or a student that is dedicated enough to see it through. Longer amounts of time could also indicate a concept that is a little too difficult to grasp or some confounding factors outside of a student's control. Regardless, the longer the student took to turn in the assessment, the more opportunities to interact with the material, the student had.

To better understand this question, the material interaction could be broken down into number of interactions (‘mean\_active’) and how long each assessment took (‘mean\_assessment\_length’).

References

Hamilton, I. (2024, May 31). 2024 Online learning Statistics. Forbes Advisor. <https://www.forbes.com/advisor/education/online-colleges/online-learning-stats/>

Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z. and Wolff, A. OU Analyse: Analysing At-Risk Students at The Open University. Learning Analytics Review, no. LAK15-1, March 2015, ISSN: 2057-7494. source: <http://archive.ics.uci.edu/ml/datasets/Open+University+Learning+Analytics+dataset>