DOES IDENTIFYING AND ADDRESSING ACADEMIC DIFFICULTIES EARLY ON CONTRIBUTE TO ENHANCED STUDENT SUCCESS AND HIGHER RETENTION RATES FOR A DISTANCE LEARNING COURSE?

Neil Anderson, Aidan McGowan, Janak Adhikari, David Cutting, Leo Galway, & Matthews Collins

School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast (United Kingdom)

Abstract

In UK universities there is a problem with academic under-performance, failure and dropout of students enrolled on programming-based courses such as computer science & software development. One way to address the issue of high dropout rates in these courses is to implement targeted interventions for students who are at risk of failing or dropping out. By providing timely interventions to students who are struggling, it is possible to improve academic performance and decrease dropout rates. This requires the ability to quickly and accurately identify these students and provide them with the support they need.

One challenge with current approaches for identifying students at risk of academic failure or dropout is that they often do not identify these students until it is too late to provide meaningful interventions. To improve the effectiveness of interventions and support for at-risk students, it may be necessary to consider additional sources of data and to implement interventions earlier in the academic process.

When working with students in a distance learning programme the problem is more complex than when working with those enrolled on campus-based programmes. The nature of distance delivery means that academic staff are often denied the opportunity to regularly observe a student's performance in a classroom or computer laboratory setting. Furthermore, the literal remoteness of a distance teaching modes often stands between an academic and a struggling student and often blocks the possibility of a quick and informal chat where the student might have outlined their academic difficulties. These are both classic examples of on-campus triggers for intervention that could help to support a student; in a distance learning setting these triggers are much less likely to happen.

Our approach to identifying students at risk of academic failure or dropout involves using a wide range of data sources, including pre-matriculation socio-demographic data, aptitude test scores, assessment results, attendance data, and Learning Management System (LMS) activity data. This diverse range of inputs can provide a more comprehensive and accurate picture of a student's academic performance and risk of struggling in their studies. We frequently recalculate the prediction of likely academic success for each student, which helps to avoid the issue of "staleness" by using the most up-to-date data available. This can help to ensure that interventions are timely and tailored to the needs of each student.

Keywords: *Academic success, retention rates, distance learning, intervention, predictive learning analysis.*

---

1. Introduction

The design and deployment of an automated system for predicting academic success and intervening with students is something that must be carefully considered. There are several areas that require special attention.

First, with the vast amount of data collected on each student at matriculation and as they progress through their academic journey, it may be tempting to use some or all of this data to trigger interventions for struggling students. However, we must consider whether it is appropriate to use data originally collected for purposes other than predicting academic success, such as demographic information collected during matriculation. We must also determine whether certain sources of data, such as aptitude test results, which may only be relevant at the start of the course, should be replaced or supplemented with fresh performance indicators like results from weekly formative assessments.

Second, we must decide on a sensible trigger for interventions. If interventions are triggered too soon, students and academics may become overwhelmed. If they are triggered too late, we may miss the opportunity to provide help and support to a student when they need it most.

Lastly, we must design a student-centred approach to intervention. While an automatically triggered alert is easy to implement, it may be viewed as impersonal by students or, worse, incite fear or worry in a struggling student. It is important that we strike the right balance between a system that is manageable and one that provides the personal encouragement a student in academic difficulties requires.

The purpose of this paper is to present an overview of the key areas that we need to take into account when designing our pilot study.

2. Understanding data sources for predicting student success

Our approach can make use of a wide range of data sources to predict academic success. We discuss five possible data sources here.

2.1. Socio-demographic data

Nawa et al. (2020) conducted a statistical analysis of the associations between demographic factors and the academic trajectories of medical students at a university in Japan. They used a multinomial logistic regression to determine the association between a student's GPA (Grade Point Average) trajectory group and demographic factors, such as high school type, high school geographical area, admission test type, high school graduation year, whether the student was a biology major, and sex. Their findings revealed that some demographic factors were associated with GPA trajectories. These factors included high school geographical area, type of admission test, high school graduation year, and sex.

While it is common for universities to collect socio-demographic data as part of their pre-matriculation process, it is unlikely that students are asked whether their demographic data can be used to predict their likelihood of academic success. Therefore, our system can only use socio-demographic data if a student has given consent for it to be used in predicting academic success.

2.2. Aptitude test scores

The utilization of aptitude testing is a widely accepted method for screening employment and academic applicants (Choi et al., 2018). The nature and extent of these assessments vary, but most focus on evaluating the candidate's capacity to quickly comprehend and analyse information. When considering the use of these tests for academic admission, the primary objective is to forecast student achievement. However, the accuracy and validity of test results are crucial to the success and sustainability of the related academic programmes, and hence, an area of substantial investigation.

Aptitude tests generally measure three fundamental competencies: Information Processing - the ability to use available information using numerical and analytical reasoning; Solution Generation - the ability to solve problems using abstract reasoning, and Decision-Making - the ability to solve problems using critical and logical thinking.

McGowan et al. (2021) found that aptitude tests, when taken as a whole, were able to provide a prediction of score outcomes on a Programming module. Furthermore, Kuncel and Hezlett (2007) determined that standardised (aptitude) tests, in conjunction with prior academic performance, were reliable indicators of success when evaluating candidates for graduate school programmes.

2.3. Assessment results

Assessment results are often used as a prompt for intervening with students who are struggling in their studies. Some academic staff may choose to reach out to students who have performed poorly on a regular assessment, such as a weekly formative quiz. However, this type of tracking can be challenging to manage consistently and fairly over time. For instance, it is time consuming to manually track each set of assessment results and then individually contact students. The process is also prone to human error. On the other hand, end-of-module assessment results can serve as a clear trigger for intervention with students who have failed or nearly failed a module. Yet, in this case, it is often too late to provide a meaningful intervention as the module has already been completed.

In our system, we will consider using the LMS API to programmatically access assessment results for every student as soon as they are posted. This will allow us to automatically contact students to offer help, advice and support.

2.4. Attendance data

Attendance data for lectures can be a useful way to indirectly measure student performance because it provides an insight into a student's engagement and participation in their course. Regular attendance is generally associated with higher levels of academic achievement, as it suggests that the student is actively involved in the learning process and is motivated to learn. Gump (2005) conducted a study on attendance data and academic success for 300 undergraduates enrolled in a general education course at a large U.S. university. As expected, the study found a strong negative correlation between absences and lower final grades.

When considering distance learning courses, the unique nature of the course delivery must be taken into account. These courses may still offer traditional live lectures, albeit through a video streaming platform like Zoom or Microsoft Teams. Hence, students may not feel the same need to attend the live session as they would in a physical classroom, particularly if a recording of the session is available. In addition, some distance learning courses offer extensive pre-recorded video content that is accessible to students before the live sessions. Due to the flexibility of distance courses, students may not feel obligated to attend a live lecture as they can acquire the same information through other means.

It should be noted that attendance should not be the sole criterion for evaluating student performance, as it is possible for a student to attend all lectures but still struggle academically, or for a student to attend very few live lectures and still complete the course successfully.

2.5. Learning Management System data

In our system design, we need to decide when to intervene. While socio-demographic data and aptitude test scores provide us with some information that can be used to predict a student's academic success, the timing of data collection - which takes place before matriculation - means that they do not necessarily reflect a student's academic progress on a week-by-week basis during the programme. Assessment results, lecture attendance data, and LMS activity data can be used to provide a "live" picture of a student's academic progress.

In a previous pilot programme, Cutting et al. (2021) used LMS data as a measure of student engagement. Their Engagement and Alerting Tool (EAT) tracked whether students interacted with learning materials and assessment pages on the LMS. If the tool detected that a student had no LMS engagement for a single module in a set period of time (7- and 14-day periods were both tested), then an automated alert was sent to the module owner. If a student had no LMS engagement with more than one module over the same period of time, then an "escalated" alert was also sent to the student's personal tutor and welfare staff. Their tool could be set to continually monitor student engagement over a full academic year. This means that their tool was able to trigger interventions using "live" data, which is much more agile than using only pre-matriculation data.

In our approach, we will consider using LMS engagement data over both 7- and 14-day periods in a similar way to the EAT pilot.

3. Student-centered interventions

In our system design, we need to decide how to intervene with a student identified as being "at risk". We considered sending an automatic email to each student identified as being "at risk" to begin the intervention; however, this approach could be seen as impersonal. CQUniversity in Australia developed a learning analytics system called Early Alert Student Indicators (EASI) (Lawson et al., 2016) to help teaching staff identify those students potentially "at risk" of failure. Their system can be considered to be semi-automated, in that while it automatically identifies which students are "at risk" of failure, a member of staff must personalize the text of the "intervention" email before it is then mail-merged and sent to the student. The use of mail-merge is important because it means the intervention email appears to come from an academic member of staff, rather than an automated email account associated with the EASI system.

Lawson et al. (2016) also undertook lexical analysis of 223,979 emails that academic staff sent to students via the EASI system. They found that the "vast majority" of emails contained customized text that could be considered positive or motivational in tone. For example, they found phrases such as: "Do you need some help?" and "Please contact me to discuss this situation, as I would very much like to help you". However, in a minority of emails, they found text that could be considered demotivational. For example: "If you do not attempt this assessment, you will fail."

A number of important lessons can be learned from the approach taken by CQUniversity with their EASI system. First, their use of mail-merge means that each student "at risk" does not receive an impersonal "alert" from EASI. Instead, each student receives a personalized email (one that they can easily reply to using their email client) from an academic member of staff. Second, unless the options for personalization of the email text are limited or very stringent rules are set for staff, it is possible that some of the intervention emails will be sent with text that is considered demotivational.

In our approach, we will consider using the Microsoft Graph API to automatically send an email intervention to each student "at risk" on behalf of a member of academic staff. In this scenario, the intervention email will be sent programmatically to the student but will appear with the sender and "reply to" specified as a member of staff. This allows the system to ensure that the email contains motivational text, and still allows the student to reply directly to the member of staff. We will also consider building in a "pre-flight" step that allows staff to prevent an email from being sent if they are already in communication with a student.

4. Discussion

We are at the very beginning of a journey, during which will ultimately seek to answer our motivating question: does identifying and addressing academic difficulties early on contribute to enhanced student success and higher retention rates for a distance learning course? In this paper, we have briefly introduced some of the key ideas and concepts that this long-term project will seek to investigate. We have identified and discussed a number of the key data sources available to us to help us measure and predict student success. We have given consideration to what a 'good' student intervention should look like.

In the short term, we plan to develop API integration software that will enable us to access the data sources outlined in this paper in a practical manner. Concurrently, we will investigate the type of student consent required for using each data source to predict academic success. Once the technical integrations are complete and necessary consents obtained, we will conduct a technical training phase during a full academic year to retrieve and store each data stream for each student. We will also record the academic outcome for each student on the programme to establish a benchmark of academic success.

In the subsequent academic year, we will conduct a testing phase that builds on the lessons learned during the technical training phase. This phase will guide us on how to make academic interventions in real-time on a per-student basis. We intend to provide in-progress results at the end of each of these phases.

References

Choi, S. Lam, S. Li K., & Wong, B. T. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society, 21*(2), 273–290.

Cutting, D., McDowell, A., Weir, E. & Trombino, G. (2021, May 27). *Beyond "Presenceism": Monitoring student engagement in the new normal* [Paper presentation]. Irish Learning Technology Association Educational Technology Conference 2021.

Gump, S. E. (2005). The Cost of Cutting Class: Attendance as a Predictor of Success. *College Teaching, 53*(1), 21-26.

Kuncel, N., & Hezlett, S. A. (2007). Assessment. Standardized Tests Predict Graduate Students' Success. *Science (American Journal for the Advancement of Science), 315*(5815), 1080-1.

Lawson, C., Beer, C., Rossi, D. et al. (2016). Identification of 'at risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Education Technology Research and Development, 64*, 957–968.

McGowan, A. Sage, P. & Adhikari J. (2021). Pre-course Prediction of Outcomes for Conversion MSC Programming Courses. *ICERI2021 Proceedings*, 2900-2908.

Nawa, N., Numasawa, M., Nakagawa, M., Sunaga, M., Fujiwara, T., Tanaka, Y., Kinoshita, A. (2020). Associations between demographic factors and the academic trajectories of medical students in Japan. *PLoS One, 15*(5), e0233371.