Using Analytics to Develop a Data-Driven Decision-Making Tool to Support Institutional Planning Towards Student Success

Colin Chibaya
Department of Computer Science, Data Science, and Information Technology
Sol Plaatje University
Kimberley

1. Introduction

The role of high education institutions cannot be over-emphasised. High-quality graduates are required to address the 21st-century challenges plaguing nations (Legrouri, 2021; Mahroeian et al., 2017). Consequently, there is an expectation for institutions to deliver on their mandate as they are increasingly supported by public funding. However, evidence suggests that institutions face numerous challenges in serving the diverse student body and meeting other stakeholders' expectations (Mahroeian & Daniel, 2021). Finding a balance necessitates having evidence-based strategies to improve decision making (Ifenthaler et al., 2021). Institutions are increasingly turning to data analytics to find insights to facilitate evidence-based decision making (Jones, 2019; Mahroeian & Daniel, 2021). This is motivated by the enormous amount of data accumulated via the numerous information systems that power institutions' academic and administrative operations (Campbell et al., 2007; Nguyen et al., 2020).

Analytics refers to the scientific process of transforming data into insights (Mahroeian et al., 2017). This view is intentionally broad. However, it is imperative to adopt a multidimensional perspective on analytics within the higher education context to effectively generate specific insights for specific goals. Common dimensions of analytics in higher education include Institutional Analytics, Learning Analytics, and Academic Analytics (Janse van Vuuren, 2020; Mahroeian et al., 2017; Nguyen et al., 2020). These different data analytics dimensions make it possible to track student journeys from pre-university entry to postgraduate and achieve other strategic goals of the university (Janse van Vuuren, 2020; Nguyen et al., 2020). This provides benefits such as management of student pipelines (e.g. identification and management of 'at-risk' students); fostering student academic success, retention and throughput rates (e.g. identification and elimination of success barriers and improving academic performance); and supporting student and curriculum development (e.g. monitoring of students engagement and behaviours, provide personalized learning, improve instructor performance) (Ifenthaler et al., 2021; Janse van Vuuren, 2020; Mahroeian et al., 2017; Nguyen et al., 2020). All these areas are instrumental to any institution's success. It is at the core of this project to use analytics to guide the decision-making of relevant stakeholders at SPU. As a secondary benefit, insights from the project can also contribute relevant knowledge on the role of analytics in higher education. Such a large-scale analytics approach would allow the opportunity to understand the student lifecycle and help various stakeholders better understand and support student journeys. In South Africa, a 2019 report by the Department of Higher Education and Training (DHET) calls on all universities to "invest in data analytics to better understand their student dropout and throughput rates and to identify productive interventions to improve the efficiency of the higher education system." (DHET, 2019).
Conducting such a large scale analytics project with longitudinal data collected over a couple of years necessitates a systematically phased approach in which insights from one phase feed into the next phase. In this regard, we have considered dividing the project into 3 phases that are purely based on the levels of study. The first phase, which is concerned with what happens during the first year of entry into university, is the primary focus of the TAU project. One issue plaguing first-year students is high dropout rates. Closely related to this is the stop out rates, entailing students that temporarily suspended their programmes. Evidence from around the world indicate that the highest dropout rates occur during the first year of study (Credé & Niehorster, 2012; Goradia & Bugarcic, 2019; van der Zanden et al., 2018). In South Africa, historical estimates suggest that the dropout rate during the first year of study could range between 50 and 60% (Nkosi, 2015).

Another critical issue is the high failure rates associated with first-year courses. Consequently, there has been an increasing interest in understanding students "at-risk". In this context, we follow the definition of “at risk" students as one having a higher likelihood of failing, stopping out, or dropping out. The "at-risk" of academic failure could include failing in a particular programme or failing any of the enrolled courses.

Equally important is understanding the factors that foster student success. However, it is imperative to acknowledge that success is quite varied and depends on the selected underlying theoretical underpinnings (van der Zanden et al., 2018). In the context of first years, we build on prior studies (Zajda & Rust, 2016) to equate success with academic performance. This view is essential to institutional stakeholders and funders as academic achievement is an indicator of programme success (van der Zanden et al., 2018). Similarly, academic achievement is also significant to students as it determines their progression to higher levels (Moss & Yeaton, 2015).

Against this backdrop of discussion, it is imperative to understand the factors determining student academic performance and students "at-risk" in the first year of study. Students at risk of dropping out, stopping out, or failing can be attributed to various factors, including socio-economic, educational, psychosocial, environmental, and institutional factors (Goradia & Bugarcic, 2019; van der Zanden et al., 2018). Understanding these factors is imperative for stakeholders to implement the necessary interventions. The theoretical underpinnings of these different factors are grounded in extant literature. Unlike traditional conceptual studies, the present study is an analytics study that treats these various factors as features in developing an analytics tool for SPU. Domain experts can then use this tool in combination with the theoretical underpinnings of each factor to better understand and apply data-driven interventions to support students.

2. Rationale

Failing, dropping-out or stopping-out of a program is a hefty loss to students, parents, and the institution. Everything possible should be done to prevent student failure, dropping out, or stopping out. There are institutional, personal, and academic factors that explain these outcomes. Learning analytics is a valuable tool to predict these outcomes and suggest countermeasures (Riestra-González et al., 2021). However, the use of analytics for the characterization of students at a specific institution might not be directly generalized to other institutions (Cerezo, 2016; van der Zanden et al., 2018). This is because the learning analytics model from one institution might not consider all possible indicators available in other institutions, limiting the generalizability of such models (Wu et al., 2021). Similarly, models developed using courses sharing a similar methodology or structure might not necessarily be course agnostic and might therefore fail to produce similar results in another setting (Riestra-González et al., 2021). Consequently, each institution needs to develop its analytics model to better understands its students and apply the necessary mechanism to facilitate their success. This study is motivated by the need for SPU to use learning analytics to support its own data-driven decision making.

3. Problem statement

Learning analytics tools can lead to improved ways of achieving institutional excellence (Riestra-González et al., 2021; Wu et al., 2021). How can institutions benefit from the enormous data generated daily? Institutions cannot rely on analytics models developed in different settings because such models are tailored not to resist external validity and generalizability challenges (Wu et al., 2021). SPU lacks a learning analytics tool that supports management in institutional decision-making. The problem addressed in this study is the development of a learning analytics tool to track students' performance and identify factors that affect student performance.

4. Aim, objectives, and research questions

This aim of this project is to develop a learning analytics tool to support data-driven decision making at SPU. Three objectives arise: (a) to identify the features vital for characterizing a student's profile and behaviours, (b) to use the identified features to create an early warning system that uses predictive modelling and counterfactual reasoning, and (c) to evaluate the learning analytics tool in identifying students at risk and aiding institutional decision making. Three questions are asked as follows: (a) what are the features that characterize a student's profile and behaviours? (b) how can the features be used to create an early warning system that uses predictive modelling and counterfactual reasoning? and (c) to what extent does the learning analytics tool accurately predict student performance and identify students at risk?

5. Method

This study is grounded in the theoretical foundations of design science research (DSR). Depending on the context, DSR is used as a research paradigm (Gregor & Hevner, 2013) or research methodology (Peffers et al., 2007). In the former, the paradigm informs the approach used in developing an artefact underscored by the process followed (Gregor & Hevner, 2013; Peffers et al., 2018; Venable et al., 2017). In the latter, a selected research paradigm guides DSR and the core knowledge contribution of the study (Peffers et al., 2007; Venable et al., 2017). In this study, the artefact is a data-driven decision support tool used to monitor student success and characterize at-risk students.

8.1 Research design

A mixed-method research design is followed. Precisely, explanatory sequential design is used, following a two-step approach that commences with a quantitative approach, followed by a qualitative approach (Creswell & Clark, 2018). Prior studies (Gregor & Hevner, 2013; Peffers et al., 2007) highlight use of various methods in demonstrating an artefact for triangulated tests (Bisandu, 2016). The experiments are aligned to the discussed epistemic foundation. The interpretability and explainability of the machine learning models developed as the primary artefact are paramount. In fact, local interpretable model-agnostic explanations (LIME) (Ribeiro et al., 2016) and Shapley values (Lundberg and Lee, 2017) are used to explain individual predictions in line with prior studies (Hall & Gill, 2019; Ribiero et al., 2016; Smith et al., 2021). Similarly, counterfactual explanations applied to individual students are generated (Molnar et al, 2020) to describe "what if" scenarios. Interpretability and explainability allow the evaluation of the usefulness of models, basing the evaluation of the tool on its usefulness to stakeholders. This step follows a qualitative approach anchored on focus groups. Focus groups are particularly important to find a shared understanding of the tool by various stakeholders. DARPA's evaluation framework for explainable artificial intelligence (Gunning & Aha, 2009) is used as the tool to guide the focus group interactions.

8.2 Data collection

Secondary data collected by SPU through various touchpoints (e.g. admissions, LMS, and other institutional systems) is used. Data captured during admission is used to define pre-entry features. Log files data from Moodle is used to create additional features for the tools. Data regarding outcome variables such as dropout, stop out, and performance is obtained from institutional information systems such as the iEnabler.

8.3 Data/results analysis

Quantitative and qualitative methods are employed owing to the study's mixed-method design. Quantitative analyses compare metrics such as the accuracy of models. Qualitative analyses include content analyses and micro-interlocutor analyses that are essential to ensure validation of explainability from stakeholders.

9 Achievements

The learning analytics tool is under development. Two scoping articles are concluded and submitted for reviewing is journals. The TAU project has taken a good shape to support upcoming institutional projects.

10 Challenges

There aren't any major challenges to mention. This research will, in fact, continue for the next three years.

References

Bisandu, D, B. (2016). Design science research methodology in Computer Science and Information Systems. In the International Journal of Information Technology.

Braunstein, A.W., Lesser, M., & Pescatrice, D.R. (2008) The impact of a program for the disadvantaged on student retention. College of Student Journal, 42 (1), 36-40

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M.P., & Núñez, J.C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. Computers & Education, 96, 42-54

Credé, M., and Niehorster, S. (2012). Adjustment to college as measured by the student adaptation to college Questionnaire: A quantitative review of its structure and relationships with correlates and consequences. Educational Psychology Review, 24.

Creswell, J.W., & Clark, V.L.P. (2018). Designing and Conducting Mixed Methods esearch. United States: Sage Publications, Inc.

DHET. (2019). 2000 To 2016 First Time Entering Undergraduate Cohort Studies For Public Higher Education Institutions: https://www.dhet.gov.za/HEMIS/2000%20TO%202016%20FIRST%20TIME%20ENTERING%20UNDERGRADUATE%20COHORT%20STUDIES%20FOR%20PUBLIC%20HEIs.pdf

Goradia, T., and Bugarcic, A. (2019). Exploration and evaluation of the tools used to identify first year at-risk students in health science courses: A systematic review. Advances in Integrative Medicine, 6(4), 143-150.

Gregor, S., & Hevner, A.R. (2013). Positioning and Presenting DSR for Maximum Impact. MIS Quarterly , 37(2), 337-355

Gunning, D., & Aha, D.W. (2009). DARPA's Explainable Artificial Intelligence Program. AI Magazine, 44 - 58

Hall, P., & Gill, N. (2019). An introduction to machine learning interpretability. O'Reilly Media, Incorporated.

Ifenthaler, D., Gibson, D., Prasse, D. et al. (2021). Putting learning back into learning analytics: actions for policy makers, researchers, and practitioners. utting learning back into learning analytics: actions for policy makers, researchers, and practitioners. Education Tech Research Dev 69, 2131–2150. https://doi-org.ufs.idm.oclc.org/ 10.1007/s11423-020-09909-8

Janse van Vuuren, E.C. (2020). Development of a contextualised data analytics framework in south african higher education: evolvement of teacher (teaching) analytics as an indispensable component. S. A. Journal of Higher Education, 34(1), 137-157.

Jones, K.M.L. (2019). Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. Int J Educ Technol High Educ 16, 24.

Legrouri A. (2021). The African Dream for Quality Higher Education. In: van't Land H., Corcoran A., Iancu DC. (eds) The Promise of Higher Education. Springer, Cham. https://doi-org.ufs.idm.oclc.org/10.1007/978-3-030-67245-4_49

Lundberg, S. M., & Lee, S. I. (2017, December). A unified approach to interpreting model predictions. In Proceedings of the 31st international conference on neural information processing systems (pp. 4768-4777).

Mahroeian, H., and Daniel, B. (2021). Is New Zealand's Higher Education Sector Ready to Employ Analytics Initiatives to Enhance its Decision-making Process? Int J Artif Intell Educ (2021).

Mahroeian, H., Daniel, B. and Butson, R. (2017). The perceptions of the meaning and value of analytics in New Zealand higher education institutions. Int J Educ Technol High Educ 14, 35

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable machine learning–a brief history, state-of-the-art and challenges. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 417-431). Springer, Cham.

Moss, B.G., and Yeaton, W.H. (2015). Failed warnings: Evaluating the impact of academic probation warning letters on student achievement. Evaluation Review, 39, 501-524.

Nguyen, A., Gardner, L., and Sheridan, D. (2020). Data Analytics in Higher Education: An Integrated View. Journal of Information Systems Education, 31(1), 61-71.

Nkosi, B. (2015). Of the 18% of matrics registered at universities, half drop out. Available at: https://mg.co.za/article/2015-05-21-18-of-matrics-register-at-universities_-half-drop-out/

Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: introduction to the special issue on exemplars and criteria for applicable DSR. European Journal of Information Systems, 27(2), 129-139.

Peffers, K., Tuunanen, T., Rothenberger, M.A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3), 45-77

Riestra-González, M., Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & Education, 163, 104108

Smith, B. I., Chimedza, C., & Bührmann, J. H. (2021). Individualised help for at-risk students using model-agnostic and counterfactual explanations. Education and Information Technologies, 1-20.

van der Zanden, P., Denessen, E., Cillessen, A., and Meijer, P.C. (2018). Domains and predictors of first-year student success: A systematic review.

Venable J., Pries-Heje J., Baskerville R. (2017) A Comprehensive Framework for Evaluation in Design Science Research. In: Peffers K., Rothenberger M., Kuechler B. (eds) Design Science Research in Information Systems. Advances in Theory and Practice. DESRIST 2012. Lecture Notes in Computer Science, vol 7286. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29863-9_31

Zajda J., Rust V. (2016). Current Research Trends in Globalisation and Neo-liberalism in Higher Education. In: Zajda J., Rust V. (eds) Globalisation and Higher Education Reforms. Globalisation, Comparative Education and Policy Research, vol 15. Springer, Cham. https://doi.org/10.1007/978-3-319-28191-9_1