Using Analytics to Uncover Early Determinants of Academic Performance for Adult Learners

Chong, S., Singapore University of Social Sciences, Singapore
Lee, Y. H., Singapore University of Social Sciences, Singapore

The IAFOR Conference for Higher Education Research- Hong Kong 2019
Official Conference Proceedings

Abstract
By and large, the arrival of the digital age have accelerated the development of analytics to guide data-informed efforts in teaching and learning. This has also transformed the way how higher education institutions look to optimize student success. In this study, through the use of data mining techniques, the university gained a better understanding of variables that influenced the adult learners first year academic performance. In particular, the results from the CHAID (or Chi-squared Automatic Interaction Detector) model highlighted the importance of previous academic performance and behavioural variables such as credit units taken and withdrawn in predicting learners at risk. The findings resonated with the opinion that an adult learner may find it challenging to juggle the demands of higher education, work-life, and family-life concurrently, at the onset. Henceforth, this group of struggling adult learners may benefit from a better management of course loading, as early as possible.

Keywords: Data mining, academic performance, data-informed efforts

Introduction

There is a growing diversity of learners pursuing higher education (McLaughlin, et. al 2013). The fastest growing population is the working adult learner (Chong, Loh & Babu, 2015). This is significant, as more and more adults who have been out of school for some years are turning to higher education institutions to start, continue or complete undergraduate or graduate degrees. These working adult learners are no longer in a traditional learning environment and they have the option of taking varied paths to degree completion (Moore & Shulock, 2009). Higher education institutions have to address this growing population to maximize their potential and retain the optimal number of students.

Higher education plays a fundamental role in creating competitive advantage for the society. While societies are increasingly dependent on skilled and talented workforce, many are facing population changes. Education in Singapore is facing challenges of rapidly changing demographics. Society is aging at the same time that the birth rate is falling (Singapore National Population and Talent Divisiion (NPTD), 2013). A key pressure felt throughout the educational system is the increasing participation rate of non-traditional students.

The ease of data collection and advances in information technologies, such as storage capability, processing power and access speed, has enabled educational institutions to accumulate vast amounts of data. A significant amount of data/analytics-driven activities has been undertaken in higher education. A key goal of analytics in education is to transform and improve teaching and learning by the use of data (Pinnell, Paulmani & Kumar, 2017). A report by a US think tank for Center for Data innovation has reproduced a paper advocating a vision for “data-driven system” for education and how through the enhanced use of data to significantly improve how educators teach and how educational administrators manage (New, 2016). The study presented in this paper, through the use of data mining techniques, helped the University gain a better understanding of the variables that influenced the adult learners academic performance.

Considerable research has pointed to the importance of identifying risk factors of higher education learners as early as possible, since early identification can lead to early intervention and increasing the likelihood of success (Upcraft, Gardner & Barefoot, 2005). Entering characteristics accounted for the greatest contribution to retention to the second year of university (Fursman, 2012). Research directs our attention to the importance of the first year of university as critical to the likelihood of undergraduate degree attainment--as the greatest attrition rate occurs before the second year of university (Adelman, 2006). With adults constituting an increasing portion of today’s student body, it is important to find out how we can better support their learning. The key direction of this study is to develop predictive model to identify early predictors of academic performance of adult learners who are enrolled in blended undergraduate programmes. This paper, a subset of the study, focuses on identifying variables to predict at-risk adult learners at the end of 2 semesters with the university.

Review of Literature

Learning analytics and data mining have emerged in the growing abilities of educational institutions to capture a rapidly increasing amount of data to "develop models for improving learning experiences and improving institutional effectiveness" (Huebner, 2013). The process is often initiated without any preconceived hypothesis, adopts a data analysis methodology (Tiwari, Singh & Vimal, 2013) and is often interchangeable with the term knowledge discovery in databases with the aim of obtaining insightful and useful findings (Giudici, 2013). Kovacic (2010) constructed prediction models on students' success based on enrolment data with statistical techniques such as CART (classification and regression technique) and QUEST (quick, unbiased and efficient statistical tree). He concluded that classifying students based on pre-enrolment data helps to identify students who may be at-risk and based on his findings he recommended orientation, advising and mentoring programs to support these students.

The literature also indicated that these learning analytics or algorithmic approaches towards predictive modelling could provide more insightful findings vis-à-vis traditional statistical modelling approaches (Li, Nsofor & Song, 2009; Bogard, James, Helbig & Huff, 2012). Amburgey and Yi (2011) says that the primary goal of data mining should be to use the data collected at colleges and universities to predict outcomes. These approaches allow the higher education institutions to analyse individual's potential for success. Such predictions are useful to identify and support students with appropriate interventions to improve their academic performance.

Methodology

In this paper three groups of variables are used – (1) student information such as gender, age, race, years of working experience, (2) academic information such as prior academic institution, prior academic performance (including O-Level English & Mathematics grades), as well as, (3) student-level data up to their 1st semester of study at the University such as their 1st semester (SEM1) GPA, proportion of credit units completed and withdrawn in their 1st semester. The approach undertaken in this study is directed by the Cross Industry Standard Process for Data Mining (or CRISP-DM framework), and is summarised in Figure 1.

The dataset comprises of 1,912 student records from the same intake year. The variable of interest (or the target variable[1]) is a derived dichotomous variable called Semester 2 at-risk based on the Semester 2 semestral grade point average (SGPA). To derive this target variable, students who obtained a Semester 2 Semestral GPA (SGPA) score of less than 2.3 is defined as '1' while those who obtained a Semester 2 GPA score of 2.3 and above is defined as '0'.

The main bulk of the time expended in the data preparation stage is on the treatment of missing values from the input data, as well as, deriving new variables such as the

---

[1] Students who have performed below 2.3 GPA (out of a 5-point grade) in their 2nd semester denoted as '1' indicating the student is at risk, otherwise '0' indicating the student is not at risk.

relevance of diploma to current degree programme. Both of which, may influence the results of the modeling.

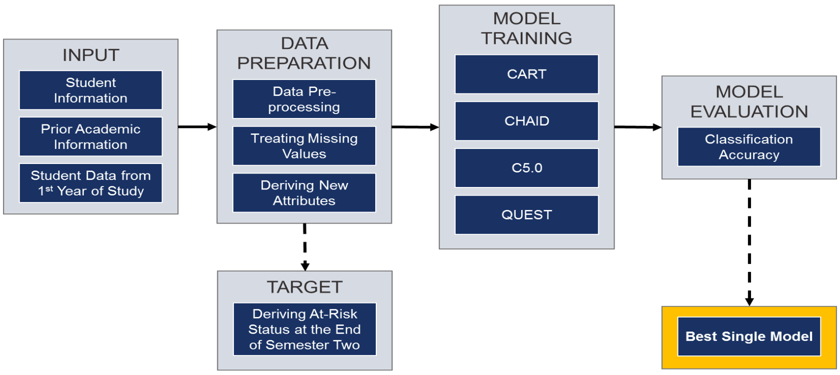

Figure 1: Predictive Modeling Process Flow

The IBM SPSS Modeler is used on the data to find a classification model to predict Semester 2 At-risk. Model development started with the building of a baseline reference model using the CHAID, Neural Network, C5.0 and CRT algorithms. Numerous decision tree-based algorithms were used in the model training stage to assess the importance of the input variables in relation to the variable of interest. Out of the few algorithms, even though the Neural Network model offers the best performance statistics, it did not offer information that explained the inner relationship between factors within the model. The C5.0 model appeared over-fitted with a complex tree that terminated with a large number of child nodes that impacted its interpretability.

The final decision tree model chosen for this study is a CHAID (Figure 2). The model was then evaluated for stability using the 10-fold cross validation method. The CHAID model with an in-sample accuracy of 74.3%, sensitivity of 69.7% and specificity of 73.4% in predicting students being at-risk at the end of Semester 2. In cases where instability is detected, the research team studied the dataset and the probability plots to identify outliers. Once outliers were identified and accounted for, a stable CHAID model was derived.

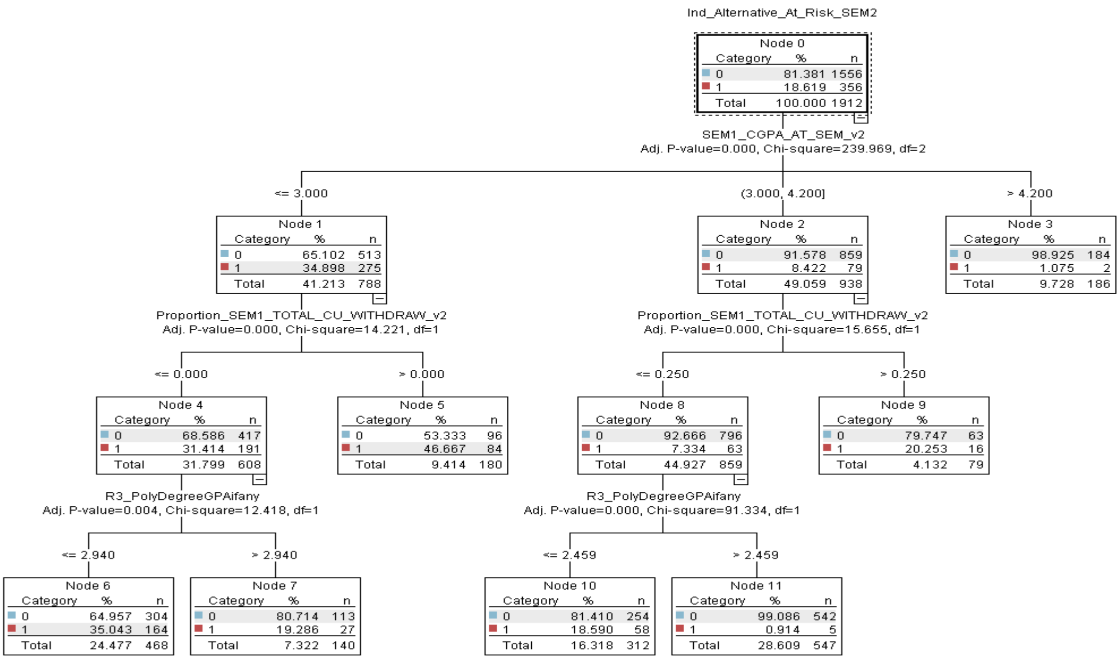

Figure 2: Decision Tree Visualisation (CHAID)

The structure of the chosen CHAID model is broadly summarised in Table 1 below.

Table 1: Summarised Structure of the Sem2 at-risk model

1st split criterion : SEM1 SGPA	≤ 3.00 (low band score)			3.00 to 4.20 (mid band score)			>4.20 (high band score)
Probability of Sem2_Outcome = at-risk	0.349			0.084			0.011
2nd split criterion : Proportion of Sem1 CU Withdrawn SGPA	≤0.00		>0.00	≤0.5	>0.25		No further split
Probability of Sem2_Outcome = at-risk	0.314		0.467	0.073	0.203		
3rd split criterion : Poly GPA	≤2.94	>2.94	No further split	≤2.46	>2.46	No further split	
Probability of Sem2_Outcome = at-risk	0.350	0.293		0.186	0.091		

Conclusion

As explained earlier, the choice of using decision tree-based algorithm was due to its explanatory power. The decision tree visualization (see Figure 2) allows end users (including programme administrators, educators, and researchers) to evaluate the impact and interaction of the variables in a more intuitive manner.

The selected decision tree shows that adult learners who had not performed so well (those scoring lower than 3.0 GPA) in their 1st semester, and who had also withdrawn credit units (CU) during their 1st semester, were more likely to be at risk in the 2nd semester. The variable Proportion of Semester 1 CUs withdrawn may be a proxy to whether the adult learner is able to cope with the study load. This might be an indication of early signs of distress in these adult learners, in managing the pace and demands of the curricula in the University. On the right side of the decision tree, adult learners who had performed much better (those scoring higher than 4.2 GPA) in their 1st semester, were observed to be highly unlikely to be at risk in their 2nd semester.

As these adult learners are part of a growing population, it is important to keep their constraints, needs, and goals in mind when examining the quality of their educational experiences. These adult learners may require different kinds of support than their traditional-aged peers in an undergraduate programme. Their needs and constraints may be completely different from those of traditional-aged students; therefore it is important to consider their experiences to ensure the support of all students' success.

This study has provided insights to understanding and facilitating adult teaching and learning. Certainly, given the nature of adult learners enrolled in the programmes, some changes to be considered may include minimising the course loading for learners who are more likely to be at risk. Academic advisement may include coping strategies as well as registering for a manageable course load, especially if the adult learner has to balance studies with work and family commitments.

Admittedly, there are other variables that can help to better explain the adult learners' performance. As such, further studies can consider the impact of nature of a job, the job size, distance to and fro workplace, family nucleus, and other measures of intellectual capacity such as reading and critical writing skills, if desired. Other decision trees and ensembles of models can also be explored as educational institutions need to employ data mining for effective decision making, efficient operations and to improve teaching and learning (Koh & Chong, 2014). No one theory or study explains all of the individual and institutional variables that contribute to student persistence and success. Tinto, one of the leading researchers on student retention, stated "despite our many years of work on this issue, there is still much we do not know and have yet to explore." (Tinto, 2012, p. 2).

A longitudinal study to track these learners can also provide further insights on variables that can support adult learning. Future research could also consider other psychosocial factors that might predict adult learners retention and attrition in higher education. Aside from enhancing the accuracy of prediction, one of the directions for future research could be focused on using the data collected to identify and develop the support systems for teaching and learning.

A natural follow up to this study is one that looks at predictors of retention of adult learners. Once enrolled, what is the University doing to retain them? Older retention studies may not be representative of today's diverse student population in higher education (Tinto, 2012), a study of persistence and access is essential as we want our adult learners to have a positive university experience and to complete their academic goals.

References

Adelman, C. (2006). *The Tool Box Revisited. Paths to degree completion from high school through college. US Department of Education*.

Chong, S., Loh, W. M., & Babu, M. (2015). The millennial learner: A new generation of adult learners in higher education. *Advances in Scholarship of Teaching and Learning*, *2*(2), 1-14.

Fursman, P. M. (2012). The National Survey of Student Engagement as a Predictor of Academic Success. Retrieved 27 December, 2017, from http://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?article=1154&context=etds

Koh, H.C., & Chong, S. (2014). *Broadening and deepening of SoTL with learning analytics.* Retrieved from 27 December, 2017, http://tlc.unisim.edu.sg/research/AdvSoTL/pdf/koh_hian_chye.pdf.

McLaughlin, P., Mills, A., Davis, P., Saha, S., Smith, P., & Hardie, M. (2013). Lifelong learning pathways: Addressing participation and diversity in higher education. Retrieved 27 December, 2017, from http://dro.deakin.edu.au/eserv/DU:30061604/mclaughlin-lifelonglearning-2013.pdf

Moore, C., & Shulock, N. (2009). *Student progress toward degree completion: Lessons from the research literature.* Sacramento: The Institute for Higher Education Leadership and Policy.

New, J. (2016). Building a data-driven education system in the United States. *Center for Data Innovation, November*, *25*.

Pinnell, C., Paulmani, G., & Kumar, V. (2017). Curricular and Learning Analytics: A Big Data Perspective. In *Big Data and Learning Analytics in Higher Education* (pp. 125-145). Springer International Publishing.

Singapore (2013) National Population and Talent Division (NPTD), Prime Minister's Office, *A Sustainable Population for a Dynamic Singapore: Population White Paper*. Retrieved on 27 December, 2017 http://www.nptd.gov.sg/content/NPTD/news/_jcr_content/par_content/download_98/file.res/population-white-paper.pdf

Tinto, V. (2012). *Completing college: Rethinking institutional action*. University of Chicago Press.

Upcraft, M., Gardner, J., & Barefoot, B. (2005). Challenge and support: Creating climates for first-year student success.

Contact email: sylviaChong@suss.edu.sg, leeyh@suss.edu.sg