DIY ANALYTICS FOR POSTSECONDARY STUDENTS

Timothy Arndt* and Angela Guercio**
**Dept. of Computer and Information Science, Cleveland State University. Cleveland, OH, USA*
***Dept. of Computer Science, Kent State University. Kent, OH, USA*

ABSTRACT

Recently organizations have begun to realize the potential value in the huge amounts of raw, constantly fluctuating data sets that they generate and, with the help of advances in storage and processing technologies, collect. This leads to the phenomenon of big data. This data may be stored in structured format in relational database systems, but may also be stored in an unstructured format. The analysis of these data sets for the discovery of meaningful patterns which can be used to make decisions is known as analytics. Analytics has been enthusiastically adopted by many colleges and universities as a tool to improve student success (by identifying situations which call for early intervention), more effectively target student recruitment efforts, best allocate institutional resources, etc. This application of analytics in higher education is often referred to as learning analytics. While students of post-secondary institutions benefit from many of these efforts, their interests do no coincide perfectly with those of the universities and colleges. In this paper we suggest that post-secondary students might benefit from the use of analytics which are not controlled by the institutions of higher learning – what we call DIY (Do It Yourself) analytics – a set of tools developed specifically to meet the needs and preferences of postsecondary students. The research presented in this paper is work in progress.

KEYWORDS

Academic analytics, student-centered learning, postsecondary education, learning analytics.

1. INTRODUCTION

Recently organizations have begun to realize the potential value in the huge amounts of raw, constantly fluctuating operational data sets that they generate and, with the help of advances in storage and processing technologies, collect in transactional systems [1]. The latest techniques from computer science, mathematics and statistics are needed to perform this analysis and generate strategic insights, e.g. about visitors to a company's website for better marketing efforts, resulting in the growing importance of the field of business analytics. Both structured data (stored in relational and non-relational database systems) and non-structured data can be analyzed using data mining techniques and the results presented using information visualization methods to best guide organizations' decision makers [2, 3, 4]. Data mining [12] is sometimes differentiated from analytics by the way that analytics tests for specific hypotheses while data mining lacks a hypothesis, instead searching large data sets for interesting patterns. Some other experts consider data mining to be a part of analytics.

Analytics has been enthusiastically adopted by many colleges and universities as a tool to improve student success (by identifying situations which call for early intervention), more effectively target student recruitment efforts, best allocate institutional resources, etc. [5, 6]. The evolution of big data and its widespread adoption in American higher education has been documented by Picciano [8]. Sources of the data which can be analyzed include institutional data about students, courses, applicants, however, a particularly rich field to mine for data is associated with online courses and Course Management Systems (CMS) [11]. Information extracted from a CMS can be quickly assessed for early warning signs of student failure, leading to prompt intervention and increased chances of student success as well as higher student engagement. Such data can also be used for student assessment and course redesign.

Academic analytics uses a combination of institutional data, statistical analysis, and predictive modeling to create insight which students, instructors, or administrators can use to develop a strategic plan for enhancing academic outcomes. The University System of Georgia carried out an experiment using analytic techniques to develop an algorithm to predict student completion and withdrawal rates in an online

environment. The results helped to confirm that it was possible to predict accurately the likelihood that a student would successfully complete an online course [6].

Goldstein [9] proposes the term "academic analytics" as an alternative to "business intelligence" in the academic realm. He surveys seven areas that analytics can be used in academia: advancement/fundraising; business and financing; budget and planning; institutional research; human resources; research administration; academic affairs. The Signals project at Purdue University has delivered early successes in academic analytics, prompting additional projects and new strategies [10]. A visual analytic tool being used for student enrollment is shown in figure 1. Clearly, most of these areas are not of interest to students in post-secondary education, except perhaps tangentially. Our DIY approach will concentrate on areas that are directly of interest to students.

College	Department	Applied	Admitted	Enrolled	Admission Rate	Apply Yield	Admit Yield
Business Administration	Accounting	561	244	46	43.49%	8.20%	18.85%
	Finance	425	177	36	41.65%	8.47%	20.34%
	Managerial Studies	1767	715	155	40.46%	8.77%	21.68%
Subtotal: Business Administration		2753	1136	237	41.26%	8.61%	20.86%
Education	Curriculum & Instruction	3720	1426	226	38.33%	6.08%	15.85%
	Educational Policy Studies	1941	752	128	38.74%	6.59%	17.02%

Figure 1. SAS Visual Analytics for student enrollment. (http://www.sas.com/software/visual-analytics/demos/student-enrollment.html)

2. DIY ANALYTICS

While institutions of higher learning have increasingly relied on learning analytics, and while the use of analytics by the institutions can be of help to students (for example, by identifying if they are at risk of failure in a course or in a program of study, and providing intervention to help them), the needs and requirements of institutions and students are not identical. As an obvious example, students have an interest in enrolling in the institution which gives them the best chance to succeed in their chosen field, however a particular institution has an interest in getting that student to enroll, even if there is some other university which would better meet the student's needs.

Consider also the following examples of divergent interests for institutions and the students enrolled in the institution. Students would like (all else being equal) to enroll in classes taught by professors that give them the best chance of achieving their goals. Institutions (colleges, departments) don't have any interest in steering students towards particular instructors and away from others. On the contrary, the department's interests are best served by having level enrollments in all sections of courses, rather than having some very large sections and other very small ones. Another example is in the choice of a field of study within an institution. The student's needs to discover the best program for him or her might not coincide exactly with those of the institution, which might want to steer students towards favored programs or away from others which might be getting too large.

In all of these cases, we propose the introduction of what we call DIY (Do It Yourself) Analytics for students. The name refers to the fact that the student himself will be using the analytics tools to make his best choices, rather than relying on the institutional filter (the DIY name is not meant to suggest that students will be creating/programming these tools themselves, only that they will be the end users). Student-centric learning analytics tools should be developed to allow students to reach their academic potential. We have identified in the above scenarios several types of insights that students would find valuable in the course of their academic careers, but these just scratch the surface. There are many more possibilities that can and should be explored. We will be expanding the list of possible topics in future research.

One issue that arises immediately is – where will the information that will be the input to the analytics process come from? In organizations, this is not a problem, since the organizations own the data that they generate. The situation is different in this case however, since the students do not generate or own the data that they need for DIY Analytics to work. The institutions of higher education could make this information publicly available so that it could be used by students (after it has been suitably scrubbed to make sure that privacy concerns are met). If the institutions are unwilling to make the information available, they may need to be encouraged to do so (by government agencies in the case of publicly-funded universities, by donors or accreditation boards in the case of private institutions). Some of this information is already publicly available (sometimes due to government regulations) either individually at the institutions, or collected by agencies or commercial entities (see figure 2, with information about American universities collected by US News and World Report). The Predictive Analytics Reporting Framework (PAR) project has shown that multiple universities can work together to unify and aggregate their data [7]. Such work provides hope for implementing DIY analytics.

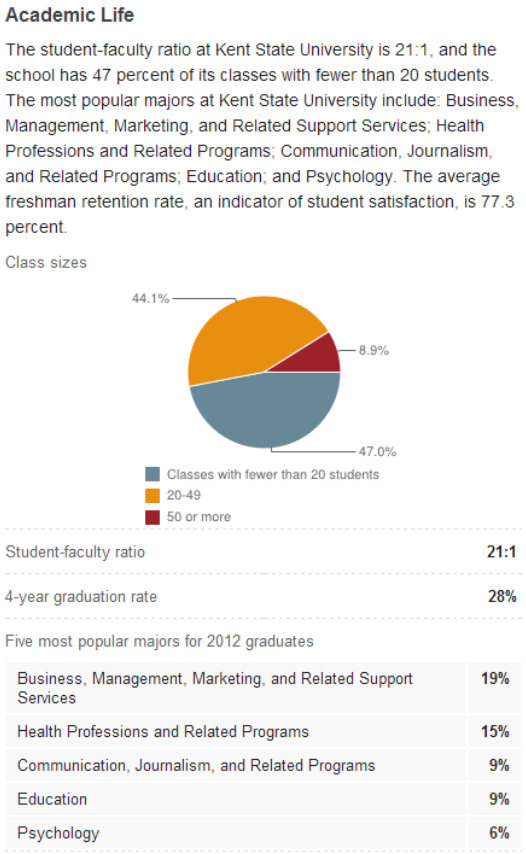

Figure 2. Information about an American institution of higher learning (http://colleges.usnews.rankingsandreviews.com/best-colleges/kent-state-university-3051)

Further initial reflection on this research indicates that in order for the information to be relevant for decision-making for non-expert users (students) this use of visual analytics will be crucial. Furthermore, given the platform preferences of today's students, the data presentation should be accessible from mobile

platforms (smartphones and tablets). This idea is reflected in our future prototype system in this area, which is described in the following section.

3. CONCLUSIONS AND FUTURE RESEARCH

This paper has described our work-in-progress in the area of DIY Analytics. We have identified several scenarios where the interests of institutions of higher education and their students diverge, leading to an opportunity to add value for students. We have also identified a possible problem in the implementation of this idea – the lack of ownership by the students of the data involved, though we hope to be able to overcome this problem in the short term by scraping publicly available data off of university websites (along with government agency and other organization sites) and in the medium and long term through a more open access to institutions' data (scrubbed for privacy). Further, we have identified a few areas that DIY Analytics must address, based on the target audience. We continue to refine all of these ideas.

We are currently in the initial stage of developing a prototype system in DIY Analytics. Our prototype will allow prospective students to explore various programs at multiple universities. The system will use SAS solutions for Hadoop [13]. SAS Visual Analytics and SAS Mobile BI will be used to produce an application accessible from mobile devices to meet the needs of today's students. Interviews with current university students will be used as part of the design process of this prototype and it will be evaluated by experiments with a group of target users. These results will be reported in a future paper.

ACKNOWLEDGEMENTS

Timothy Arndt would like to acknowledge the help of Cleveland State University MCIS students Haci Karahasanoglu and Kathleen Justice in the development of some of the ideas in this research.

REFERENCES

1. Kohavi, R., Rothleder, N. J., and Simoudis, E. 2002. Emerging trends in business analytics. Commun. ACM 45, 8 (August 2002), 45-48.
2. Naone, E., "The New Big Data". Technology Review, MIT, http://www.technologyreview.com/news/425090/the-new-big-data/.
3. "Big Data: The next frontier for innovation, competition and productivity as reported in Building with Big Data", The Economist. May 26, 2011, http://www.economist.com/node/18741392.
4. LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N., "Big data, analytics and the path from insights to value", MIT Sloan Management Review, 21 (2014).
5. Mattingly, K. D., Rice, M. C. and Berge, Z. L., "Learning analytics as a tool for closing the assessment loop in higher education", Knowledge Management & E-Learning: An International Journal, vol. 4, no. 3, (2012): 236-247.
6. Baepler P., Murdoch, C. J., "Academic Analytics and Data Mining in Higher Education", International Journal for the Scholarship of Teaching and Learning, 4 (2010).
7. Ice, P. et al., "The PAR Framework Proof of Concept: Initial Findings from a Multi-Institutional Analysis of Federated Postsecondary Data", Journal of Asynchronous Learning Networks, 16 (2012): 63-86.
8. Picciano, A. G., "The Evolution of Big Data and Learning Analytics in American Higher Education", Journal of Asynchronous Learning Networks, 16 (2012): 9-20.
9. Goldstein, P. J., and Katz, R. N., "Academic Analytics: The uses of Management Information and Technology in Higher Education", EDUCAUSE Center for Applied Research (2005).
10. Arnold, K.E., "Signals: Applying Academic Analytics", Educause Quarterly, 33 (2010).
11. Romero, C., Ventura, S., Garcia, E., "Data Mining in Course Management Systems: Moodle Case Study and Tutorial", Computers & Education, 51 (2008): 368-384.
12. Romero, C., Ventura, S., "Educational Data Mining: A Survey from 1995 to 2005", Expert Systems with Applications, 33 (2007): 135-146.
13. SAS Solutions for Hadoop http://www.sas.com/en_us/software/sas-hadoop.html