Building an Academic Analytics Capability at KSU

Han Reichgelt[1]

Executive Summary

Academic Analytics is the application in academia of techniques that have used successfully by businesses to enhance their operational and strategic decisions. Essentially, the techniques involve the discovery of trends in the data that institutions have collected to support their current operations to predict the future.

At a time where universities are held to ever growing standards of accountability and are asked to do so with dwindling resources, academic analytics promises to be an extremely useful set of techniques to help universities enhance student success, as well as the achievement of their other goals.

KSU is well placed to become a leader in the use of academic analytics both within the University System of Georgia and nation-wide. It also meets many of the conditions that a successful university-wide implementation of these techniques would require, including accurate data, an interest in the techniques at the higher management levels in the organization, and the presence of an appropriate technology platform. What is lacking is a wide-spread awareness of the power of academic analytics and the academic analytic tools that KSU already has available, and a wide skill basis to use the tools. However, given this analysis, it is not hard to see how the institution should proceed if it wishes to enhance its use of academic analytics.

[1] I am grateful for KSU to give me the opportunity to do this project and the many people within KSU who generously donated their time to help me complete the project. They include Lynn Black, Erik Bowe, Kenneth Bridges, Angie Conti, Leigh Funk, Donna Hutcheson, Teresa Joyce, Thomas Kolenko, Victor Marshall, Thomas Pusateri, Ed Rugg, Val Whittlesey, and many others. I am most grateful to each of them.

Introduction

Recognizing the need to encourage and support leadership development throughout the University System of Georgia, the System recently started an Executive Leadership Institute (ELI) to prepare selected individuals within the System to prepare for potential advancement within the System (for details, see http://www.usg.edu/executive_leadership_institute/). ELI scholars are expected to shadow a senior executive within the system, and to observe and analyze a program or process and provide recommendations to improve the efficiency or effectiveness of the program/process.

I have been fortunate in having been given the opportunity to shadow Dr Lendley Black, Provost at Kennesaw State University. After discussion with Dr Black, Associate Provost Teresa Joyce and Assistant Vice President for Academic Affairs Dr Valerie Whittlesey, we decided that the process I would analyze is KSU's capability of making more data driven decisions, also known as Academic Analytics.

This document describes the concept of academic analytics in more detail, including some examples of how other universities have used these techniques to enhance their operations, and makes some recommendations on measures KSU may wish to implement to help it improve its academic analytic capabilities. Since KSU is already well-positioned, none of these measures appear to be overly complex.

The Need for Data Driven Decision Making in Academia

Academia is under pressure from a number of different directions. For example, different agencies, such as accreditation agencies and both federal and state government, are increasingly holding universities accountable for student success. At the same time, state support for public universities is declining and public universities are put in a position in which they are asked to do more with less.

Given this situation, universities in general and public universities in particular have to become more circumspect in how they spend their resources. For example, there are a number of well-established mechanisms that universities can use to improve their retention and graduation rates, such as intrusive advising, the implementation of early warning systems, and the creation of learning communities. However, each of these mechanisms is expensive to implement and with limited funding, it is unlikely that cash-strapped universities will be able to implement all or even many of the mechanisms that have proven successful elsewhere. Moreover, not all students need these mechanisms to be academically successful. Even those universities with low graduation rates eventually graduate some of their students. The question therefore becomes how a university that wishes to improve its graduation rate can use its limited resources to help those students who are at the greatest risk of not graduating.

Fortunately, universities collect a wealth of so-called transactional data, electronic data directly related to their operations. For example, universities have extensive student records, which typically include data on a variety of demographic variables (e.g., age, gender, country and date of birth, and so on), academic data (e.g., high-school GPA, GPA at the institution, performance in individual courses, and so on), as well as, increasingly, data on extra-curricular activities. They also have significant amounts of data on their faculty and staff. Moreover, many universities have data on their facilities, such as the size of lecture rooms, room utilization, and so on.

What universities are in general less good at is using the transactional data that they collect to guide their decision making processes. The vast amounts of transactional data often contain hidden information, which, if made explicit, could help an institution deploy its resources more effectively. For example, a careful analysis of student data might identify factors that are highly correlated with academic success. It is probably unnecessary to direct resources at students who display most of these factors, and a more prudent use of resources would be to direct them to those students who past data suggest are at greater risk of academic failure.

Academic Analytics

The analysis of transactional data to help make decisions is well established in business, where the concept is known as "business intelligence". A widely used example is market basket analysis, which is used by retail organizations to determine which items are frequently bought at the same time. Retailers use this information in the design of the lay-out of the store. An interesting example is the finding that shoppers who buy disposable diapers often also buy beer (shoppers at the world's largest retailer will now understand the reason that the baby department is very close to the beer aisle). Retailers also use business intelligence to predict sales patterns depending on weather or other seasonal fluctuations. For example, the same large retailer knows that sales of alcohol go up prior to a heavy storm, while the sales of generators increase in the aftermath. Clearly, this information can be used in making decisions on how to stock and restock its retail outlets.

Since the term "business intelligence" does not necessarily sit well with academia, a number of authors (e.g., Campbell et al, 2007; Goldstein, 2005) have coined the phrase "academic analytics" to describe the application of business intelligence to academia. Campbell et al (2007) give a number of examples of the use of academic analytics in universities. For example, universities have used academic analytics to determine which students in an applicant pool are most likely to enroll in the institution, to predict and improve student retention, and to have an earlier identification of at-risk students.

There are essentially two different approaches that can be used in academic analytics. One set of techniques involves querying the transactional data to determine the validity of a user-generated hypothesis. For example, an academic adviser may believe that students who graduated from a particular high school are more likely to be successful in an introductory mathematics course than

students from another high school, and he or she may wish to query the data to determine whether this hypothesis is valid.

While user-generated hypothesis testing can be of great benefit to any organization the drawback of this approach is that it is limited by the imagination of the user. Users come to the data with certain preconceived ideas about what is likely to be true and are unlikely to explore possibilities that do not fit their conception of the world. Partly in response to this, the computer science community has developed a set of techniques to get the computer itself to generate and test hypotheses. These techniques are typically referred to as "data mining". Data mining involves a series of statistically based techniques to generate and validate interesting hypotheses. Hypotheses are deemed interesting if, among other factors, they are concise, general in the sense that they cover a non-trivial subset of the data, and reliable in the sense that they are in general true of the data (for a technical overview of how this concept is defined in the data mining literature, see Geng and Hamilton, 2006).

While it is in general fairly straightforward to provide users with the ability to test their own hypotheses, the implementation of data mining requires more powerful computing resources. However, as the beer-and-diapers example at the start of this section illustrates, the use of data mining may lead to the discovery of patterns that were unexpected. Moreover, automated data mining systems can test a far greater number of hypotheses than a human user can be expected to do.

Requirements for Implementing an Academic Analytics Capacity

There are a number of requirements that an organization would have to have in place to implement an academic analytics capability[2], namely

1. Accurate transactional data;
2. A flexible technology platform that is available to collect, mine, and analyze data;
3. A sponsor at the presidential or the vice-presidential level;
4. An organizational culture that encourages data driven decision making;
5. Staff who are skilled in data analysis and in the use of the tools that the organization has available for data mining.

KSU's Readiness

In order to determine how well KSU is positioned to enhance its academic analytics capability, it is useful to analyze the extent to which it meets each of the requirements listed in the previous section

[2] This section is partly based on Campbell et al (2007).

Data Quality: The unit within KSU whose primary responsibility is directly relevant to the establishment of enhanced academic analytics capabilities is Enterprise Information Management (EIM). EIM has implemented a virtual data center, details of which are available at http://vic.kennesaw.edu.

EIM recently established the post of Director of Data Quality Initiatives, whose primary responsibility is to ensure the accuracy of the data collected in EIM. The holder of this position, Dr Leigh Funk, along with the CIO, Dr Randy Hinds has formulated and implemented a policy on Data and Metadata Governance Structure at KSU, which closely follows the business processes established by the Board of Regents (Chapter 12 of the Board of Regents' Business Process Manual). EIM concerns over the need to improve data quality and its proper management have not only led KSU to establish this key role, but also led the University System of Georgia's Central Office to recently establish and employ a Chief Data Officer with similar oversight responsibilities for improving system-wide data quality and management. The quality of KSU's transactional and census data stores are in reasonably good shape, but substantive efforts are underway to improve data integrity.

Technology Platform: In order to perform academic analytics an institution must have in place both systems to capture transactional data and systems to perform academic analytics. In line with most other universities, KSU has good relational database systems in place to capture transactional student and personnel data. It uses Banner and PeopleSoft systems for these functions. However, while those systems perform various automated operations reasonably well (e.g. , admissions processing, registration, accounting, budgeting etc.), they are not well suited to provide easy access to the data needed for academic analytics.

While a number of universities within the University System of Georgia have systems in place to allow for the testing of user-generated hypotheses, EIM has gone one step beyond. In order to further expand KSU's academic analytic capability, KSU's CIO established EIM in 2007 and acquired a well-known business intelligence data integration platform and suite of analytics tools, developed by SAS. EIM will soon complete the integration of transactional data from KSU's Banner. PeopleSoft and other operational databases as part of constructing a SAS-powered data warehouse. These SAS solutions have many attractive features, including the fact that it is accessible over the web, the fact that it allows both for the testing of user-generated hypotheses and data mining, and the fact that it allows sophisticated users to link special purpose repositories of data with the data that EIM has loaded into SAS. All of these factors lead to the conclusion that KSU has the appropriate technology platform in place to implement a significant academic analytic capability. It should be noted that Erik Bowe's work as KSU's chief SAS Initiative architect has just been spotlighted on the corporate SAS web site as a model of innovation in higher education 's use of business intelligence and data base integration.

EIM also generates a large number of pre-defined SAS-powered routine reports, available on its web site. The EIM web site gives access to a multitude of reports regarding students, courses, degrees, faculty, finances and so on. Such reports are often used to demonstrate compliance with the requirements of external agencies (SACS and other accreditation agencies, the Board of Regents, etc)

Sponsor: It is well-established that large scale IT initiatives will succeed only if they have a sponsor at a high level within the organization. While I simply cannot comment on the attitude of the president or the other VPs at Kennesaw towards data-driven decision making, it is clear that KSU's Vice President for

Operations and Chief Information Officer, Dr Randy Hinds, has made technical support for academic analytics a major strategic initiative in his division for the past three years. The work of EIM and the SAS Initiative reflects that high-level sponsorship. The fact that the Provost and VP for Academic Affairs, Dr Lendley Black, suggested that I make academic analytics the focus of my ELI project suggests that he is aware of its importance within the ranks of academic administrators and the faculty as well. One can therefore safely conclude that Dr Black could be asked to sponsor projects to enhance KSU's use of academic analytics to improve planning, operations and decision-making in Academic Affairs, which is the heart of the university enterprise.

Organizational Culture: While having senior level sponsors is critical, it is not the only organizational factor that is relevant to whether an IT initiative will succeed. For an IT initiative to be successful, it must also mesh well with the culture of the organization. Essentially, an IT initiative is more likely to be successful if its objectives, goals and methods are congruent with those of the organization. This is even more relevant in the context of a university where shared governance implies that senior management has less ability to dictate certain initiatives to the organization than it has in other settings.

For the purposes of this project, we will define organizational culture as the awareness of individuals within KSU for the potential of academic analytics, the awareness of individuals within KSU of the tools, and the willingness to use these tools.

While I have not been able to conduct an in-depth survey of KSU organizational culture, discussions with members of EIM suggest that the picture is mixed. For example, there are pockets in some colleges (Coles College of Business, the Bagwell College of Education, and the College of Science and Mathematics) that appear to be fully aware of the power of academic analytics and have either developed the ability within the dean's office to analyze relevant data or have interacted with EIM on a regular basis to have queries answered. On the other hand, other colleges either are not aware of the potential of academic analytics or have decided not to use the available tools.

Staff Skilled at Data Analysis: EIM distinguishes between four levels of users with a potential interest in using the various academic analytics capabilities, namely

1. The General Public
2. SAS Data Users
3. SAS Intelligence Specialists
4. SAS Power Users

As any university, KSU can expect there to be members of the general public with an interest in data about the institution and its operations. Examples may include prospective students and their parents, individuals who may be interested in any employment opportunities at KSU, and so on. The various pre-defined reports that are posted on for example the EIM web site will meet the needs of this group of users. EIM has published its pyramid of SAS self-service data users identifying four tiers.

SAS data users are typically KSU employers who, in addition to the reports that are accessible to the general public, also need access to password-protected pre-defined reports on the EIM site. Users in this category do not need anything beyond information on where to find the various pre-defined reports and basic web surfing skills.

SAS intelligence specialists are users who access data to generate special purpose reports that are relevant to this particular area of the organization and that are based on data that are derived from

the standard transactional information sources, such as Banner and PeopleSoft. SAS intelligence specialists will typically use a SAS toolkit SAS called the "SAS Web Report Studio", and will obviously have to have the skill to use this tool.

SAS power users, finally, are users who generate special purpose reports using not only data derived from the standard transactional information sources but also from special purpose sources, for example data bases that they have generated or data collected from outside sources. SAS power users use the SAS enterprise guide as well as the SAS enterprise miner, and will obviously need the skills to do so.

While all KSU personnel are likely to have the required skills to retrieve information either as a member of the general public or as a SAS data user, there are currently very few KSU personnel members who have the skills to use the SAS Web Report Studio, the SAS Enterprise Guide or the SAS Enterprise Miner. Colleges that have used the academic analytics ability that EIM provides are most likely to have faculty who know how to use the SAS web report studio. Several applied statistics faculty in the College of Science and Mathematics are versed in the use of SAS Enterprise Guide and the SAS Enterprise Miner. However, it is clear that the skills to use these more advanced tools are not widespread. It should be noted that Dr Jennifer Priestley's use of these tools in her classes and the creation of a degree-credit SAS Certificate program at KSU is spotlighted as a Customer Success story on the corporate SAS web site as a model of educational innovation in the use of analytics.

Next Steps

The analysis in the previous section suggests that in order to ensure a wide-spread use of academic analytics at KSU a number of steps are necessary:

1. An awareness campaign across the institution of the need for more data-driven decision making and the fact that KSU has the tools available for supporting this style of decision making. While the awareness campaign can be conducted by EIM in conjunction with representatives from those colleges that are already using SAS, it is useful if this campaign is seen as having the full support from the Provost's office.
2. An identification of selected personnel in various campus offices to be trained in the use of the more advanced SAS tools. The initial roll-out should probably be restricted to the Provost's office, the various colleges and any other agencies directly reporting to the Provost but based on the experience with this initial roll-out, subsequent phases may involve other offices on campus. As part of this exercise, the institution also needs to develop additional training material. While there is training material on the EIM web site, at least some of this material appears not to be at the right level in the sense that it is either too high level or in the sense that it presupposes prior knowledge that is unlikely to be widespread among KSU personnel (or indeed the academic personnel at any university). It is likely that CETL has an important role to play in this from a faculty development perspective.
3. The establishment of an independent process to monitor the use of academic analytics across KSU. While EIM might appear to be the appropriate unit to implement this process, it has to be

recognized that EIM is also the main driver behind the initiative to improve the use of academic analytics within KSU. There therefore is an argument that the Data Trustee from Academic Affairs be charged by the Provost to monitor and facilitate the expansion of academic analytics across the institution's educational enterprise.

Conclusion

Academic analytics has the potential to make universities, including KSU, more effective and efficient in many of their academic operations. KSU is well positioned to become a major user of this technology in that it already meets many of the conditions that have to be in place to successfully implement the technology. Through the implementation of an awareness campaign and a training program to teach selected users how to use the technology, KSU could become one of the leaders with the University System of Georgia, and indeed among universities nation-wide, in the field of academic analytics.

References

1. Campbell, J., P DeBlois & D. Oblinger (2007) "Academic Analytics: A New Tool for a New Era." *Educause Review,* 42(4): 40–57.
2. Goldstein, P. & R. Katz (2005) "Academic Analytics: The Uses of Management Information and Technology in Higher Education," *EDUCAUSE Center for Applied Research (ECAR) Study,* vol. 8, key findings publicly available: http://www.educause.edu/ir/library/pdf/ecar_so/ers/ers0508/EKF0508.pdf
3. Geng, L. & J. Hamilton (2006) "Interesting Measures for Data Mining: A Survey" *ACM Computing Surveys*, 38 (3): Article 9.