

CERTAIN INVESTIGATIONS ON BIG DATA APPROACHES IN EDUCATION AND LEARNING ANALYTICS

Velmurugan, Kannaiya Raja, Saravanan Marimuthu Raja

Dept. of Computer Science, Ambo University, ETHIOPIA

ABSTRACT

The applications of computerized tools for learning management systems in education have been increasing in the last few years. Online contents by means of mobile phones, primarily smart phones enable the students to keep the learning processes around the clock. The enormous amount of data captured from student's online activities is wasted as traditional learning analytics are not capable of processing them. So in this paper it is proposed to use Big Data technologies and tools into education for capturing learning process of students, strategic operational decisions at institutions. The computing platform and teaching methods to accommodate emerging technologies for related courses are also investigated in this paper. In addition, visualization-based data discovery tools are focused on the front end of big data on helping institutions explore the data more easily and understand it more fully.

Published on: 24th– August-2016

KEY WORDS

Big Data, Business intelligence (BI), Data warehouse, Enterprise resource planning (ERP), Visualization-based data discovery

INTRODUCTION

Higher education institutions are collecting and have access to more data than ever before and data-based decisions drive institutional effectiveness. The age of big data has come for higher education as IT becomes increasingly embedded in the process that comprises 'going to college,' such as course enrolment, classroom tutoring, and student services. Also, data about student successes and failures can be had to improve both individual and cooperative outcomes across all higher education in useful ways. Data warehouses and the cloud make it possible to collect and maintain massive records on students, alumni, operations, pedagogical impact, competition, marketplace and more. Today's sophisticated analytics technology makes it easier than ever to sift through and find meaningful patterns in all that data. As a result, demand for evidence to guide and support decision making is on the rise. Visualization-based data discovery tools allow institutions to mix up disparate data sources to get custom analytical views without rigidity and easy usage that was not available earlier. Advanced analytics are used to create interactive, animated graphics on desktops, and on mobile devices and laptops. End users can see the graphics on either the same devices, or on mobile devices such as tablets, smartphones etc.

LITERATURE REVIEW

Big Data enables to mine learning information for insights regarding student performance and learning approaches from the institution database. Rather than rely on periodic test performance, [1] instructors can evaluate what students know and what methods are most effective for each student. By focusing on data analytics, professors can study learning in far better ways. Online tools enable evaluation [2] of a much wider range of student actions, such as how long they devote to readings, where they get electronic resources, and how quickly they master key concepts.

Computerized learning modules help to assess students in systematic, real-time ways. Data analytic software gives feedback about academic performance to students and professors. It also helps to predict students who need extra help or needing more hard tests. It also identifies scholastic approaches that are effective with specific students [3].

Employing software helps teachers to find out how the students learn statistics, chemistry, mathematical principles and experimental designs. This is done through pre-test and post-test evaluation by embedded assessment [4].

Beck and Mostow [5] used intelligent tutor software to comprehend whether a pupil learnt words better when re-reading an old story or a new story. Results showed on basis of reading time, mistakes, help requests and word knowledge that re-reading a story leads to half as much learning as to reading a new story.

School systems give more importance to feedback got to increase the learning process. Measurement of time spent on a specific test, skills realised and concepts mastered are used to get the feedback which is embedded in the process, so that real-time results is utilized to find what is learnt and to observe performance over time. Also, computer can alter test questions based on how the students answer earlier questions.

Kellen et al., [6] describes SAP's HANA, a Big Data analysis tool in University of Kentucky. By monitoring and evaluating the student's background data, the system calculates a "K-Score" for each student. This score shows the involvement of students in learning activities. A low score signifies an underperforming student whose needs should be taken care of.

USING ANALYTICS TO DRIVE BETTER INSTITUTIONAL PRACTICES

An educational office relies on financial, student and other institutional data to report on and manage processes. Therefore, it's a natural shift to extend that kind of data-based decision making to all areas of college or university, including the most critical area for success in today's globally competitive environment enhancing the student experience. By utilizing information from the whole campus, linking it with outside data, and using technology to determine the "right data" to analyze and utilize, a school can easily begin asking (and answering) important questions like:

- Based upon projections from degree plans, how prepared is our institution for future classes, faculty, facilities, and other needs?
- What are the programs we should offer or drop to keep up with demand and expand our value to our constituents?
- How can financial statements from previous years help us predict future budgets?
- What is the actual return on investment for our advancement events and appeals?
- What is being said about our school on social media and what are we adding to that conversation?
- How do our continuing education, workforce development, online education, and/or schools of extension stack up against the competition?
- Based on our institution-level mission and our students' personal attainment goals, what are the best metrics to predict student success at our school?

The good news is that institutions already have a big pool of data to work with. The not so great news is, "although considerable amounts of data are being collected and stored (by higher education institutions), the data is not being used effectively to make predictions or trigger proactive responses". So, let's start small with big data. What do you have to work with that you can easily access? Most institutions find that grouping data into two main categories student data and institutional data helps clarify the kinds of big, strategic questions it can address. In the first bucket, if there are like most schools, have ample data on prospects, students, and alumni it can be posted with today's interconnected and social-media driven world, institutions have a lot of information about what they are doing, saying, thinking, and buying. The second large bucket of data that have revolves around institution. Whether institutions use an enterprise resource planning (ERP) system to connect systems across campus or are still operating in the "silo" world of disparate systems that have ample information on what institution is up to and how effective it is. That includes information and data collected from and reported to the local and federal government. Examples of data that have access to right now include:

- Data on your prospects, students, and alumni
 - + Demographics
 - + SATs, GPAs, transcripts
 - + Course selection, registration, add/drop
 - + Purchased/returned text books, library activity
 - + Financial aid applications, employment to support education
 - + Financials, fees, expenses (e.g., cafeteria)
 - + Online courses (data on how your students learn)

- + Social media
- + Internships
- + Alumni connections and post-CE job placements
- + Donations
- Data on institution
 - + Retention rates, graduation rates, transfer rates
 - + Marketing statistics
 - + Enrollment, yield rate
 - + Advancement effort progress
 - + Tenured faculty and adjunct
 - + Classroom allocation
 - + Instructional design effectiveness
 - + Financials
 - + Competitive information
 - + National Center for Education Statistics, IPEDS

Now that after recognizing the abundance of information the next step is collecting and organizing data with a business intelligence (BI) analytic tool that provides a clear view of what institutions are working with. Analytic tools offer common metrics for all datasets and give the needed dashboards, reports, visualization options, and real-time monitoring that brings data to life by making it understandable. With this new knowledge and ability institution can empower every area of an institution to make better business decisions and achieve optimal performance. Some example areas where applying the basic principles of BI analytics to big data, even minimally, can have profound impact. They are:

Improve institutional operations. In an age when all of higher education is being asked to do more with less, analytics can help reduce costs by providing information needed for streamlining and refining business processes. “Many colleges and universities have confirmed that analytics can help considerably advance an institution in areas as resource allocation, student success and finance.” Powerful analytic tools allow you to study patterns of performance over time, from one semester to another or from one year to another.

Enhance pedagogy and learning. Computerized testing, tablets and other mobile devices, online learning, course management/learning management systems, and other educational technologies are giving rise to a new era of learning analytics. Using real- or near-real-time monitoring of student activity such as postings on discussion boards, class material downloads assessment results, wiki activity, and the many other transactions per student per course—faculty can more easily create optimal learning environments and continually refine pedagogies along the way. In addition, data on faculty productivity can help drive positive learning outcomes. “Measuring faculty productivity is understandably a sensitive and controversial topic, but there is increasing acceptance that it is essential to sustainability. The issue here is largely cultural, rather than technology. However, the most successful institutions will employ technology to track, manage, measure, and improve faculty productivity.

Increase student success. Many colleges and universities now employ predictive analytics to improve their student success and retention rates. Best practices call for analyzing three years’ worth of historical data to find the risk factors and positive factors that inhibit or promote student success at an individual institution. Integrating data from multiple sources also improves at-risk student intervention efforts. These risk factors are different for every institution and will only be revealed by analyzing the existing data.

In addition, Big Data in online learning space will help institutions to improve learning outcomes for distinct students. By planning a curriculum that gathers data at every step of student learning process, universities can help student needs with customized modules, learning trees in the curriculum, assignments, and feedback, which will promote improved and richer learning.

Applying analytics to big data is quickly becoming imperative for successful higher education institutions. In higher education institutions Big Data may not solve all the issues and decisions, but they play an integral role in administrative and instructional functions. Even still, change is sometimes slow at many institutions. Using big data for impactful analysis on the campus can help with that problem, too. A recent EDUCAUSE study reported that “many study participants provided examples of how analytics programs can improve processes such as communication and decision making while increasing morale. Analytics programs can foster communication

between executive leadership, IR, and IT. Institutions should not wait for a cultural shift to be fully in place before beginning an analytics program. Initiating an analytics program may help establish that culture. Facing unprecedented demands for accountability, efficiency, and effectiveness, modern colleges and universities need to use big data to identify and evaluate strategies for improving the student experience, ensuring institutional success in every area from recruitment to alumni fundraising and everything in between. Higher education institutions that leverage the power of the large quantities of data at their disposal are better equipped to make impactful data-based decisions and thrive in today's fast-moving and competitive higher education world.

There are, however, no established best practices in higher education for what to measure or which measurement methodologies produce the most meaningful results. Schools are still learning how to harvest the power of the large quantities of data sets they collect daily. This paper provides guidelines [Figure -1] and best practices for making impactful data-based decisions and identifies mistakes to avoid when handling big data on the campus.

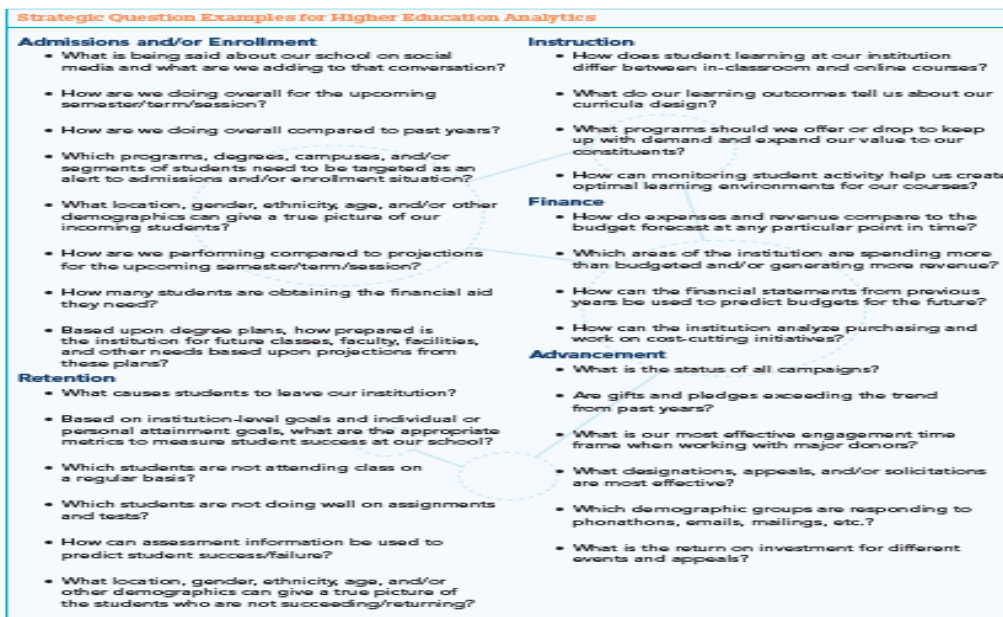


Fig. 1: Guidelines and best practices for making impactful data-based decisions

TOOLS AND TECHNIQUES

Techniques: Various techniques are utilized for the problems faces in Big Data processing. Some of the techniques utilized in educational data mining are:

- Regression – It is used in predicting values of a dependant variable by appraising the relationship among variables using statistical analysis
- Nearest Neighbor – Here, the values are predicted based on the predicted values of the records that are nearest to the record that needs to be predicted.
- Clustering – This involves grouping of records that are alike by identifying the distance between them in an n-dimensional space where n is the number of variables.
- Classification – This is the identification of the category or class to which a value belongs to, on the basis of previously categorized values.

Open Source Tools: Several Open source tools exist which help in taming Big Data. Some of top tools are:

- MongoDB is a cross platform document oriented database management system. It uses JSON like documents instead of table based architecture.
- Hadoop is a structure that allows distributed processing of big datasets across clusters of networked computers using simple programming models.

- MapReduce is a programming model and framework used by hadoop. It enables processing huge amount of data in parallel on large clusters of compute nodes.
- Orange is a python based tool for processing and mining big data. It has an easy to use interface with drag and drop functionalities with variety of add-ons.
- Weka is a java based tool for processing large amount of data. It has a vast selection of algorithms that can be used in mining data.

APPLICATIONS IN LEARNING

Big Data techniques can be utilized in different ways in learning analytics as listed:

- Performance Prediction: Student's performance can be predicted by analyzing student's interaction in a learning environment with other students and teachers.
- Attrition Risk Detection: By analyzing the student's behavior, risk of students dropping out from courses can be detected and measures can be implemented in the beginning of the course to retain students.
- Data Visualization: Reports on educational data become more and more complex as educational data grow in size. Data can be visualized using data visualization techniques to easily identify the trends and relations in the data just by looking on the visual reports.
- Intelligent feedback: Learning systems can provide intelligent and immediate feedback to students in reponse to their inputs which will improve student interaction and performance.
- Course Recommendation: This helps the student to find new courses based on their interest. This is done by analysing their activities.
- Student skill estimation: Estimation of the skills acquired by the student.

Behavior Detection: Detection of student behaviors in community based activities or games which help in developing a student model.

BIG DATA VISUALIZATION

Apache Hadoop and other technologies are used to support storage and processing, and visualization-based data discovery tools help educational institutions explore the data more easily and understand it more fully [7].

Self-service BI also assists businesses to take advantage of mobile workforces. For instance, remote and on-site members of a product development team can easily view and share visualizations that explore potential product defects or customer preferences. This bring-your-own-device (BYOD) trend means that these users can use their own mobile devices to easily explore the data, discover trends and patterns, and communicate their findings to fellow team members and other audiences.

PREDICTIVE ASSESSMENTS

Predictive and diagnostic assessments are some of the other ways of learning. In predictive assessment evaluate how students will achieve on standardized tests and diagnostic assessment highlights which techniques work for specific students and the better way to modify learning. Digital evaluation's main virtue is it gives students information needed for their learning and performance.

Performance is the key word for online predictive assessment. McGraw-Hill's Acuity Predictive Assessments [8] tool gives an initial indication of how students will perform on state NCLB assessments. It finds what the students know and what the students need to know on tests and recommends what the student should focus on to better their tests.

Similarly, the assessment tool helps "teachers probe student understanding of state standards, grade-level expectations, and specific skills, and quickly diagnose their strengths and instructional needs." By following how students solve problems and evaluate information, the tool provides guidance concerning preferred learning styles and works instruction to that preference.

Research has found that some pupils like to go through problem-solving step-by-step in a linear manner. Some others favour visual or graphical presentation in a non-linear fashion. So, assessment of learning styles is vital to personalization and tailoring instructional presentation. Digital tools that assist parents and teachers to comprehend student learning approaches are vital to educational attainment. Fifteen variables such as the number of discussion messages posted, time online, visits to course chat area, number of emails sent, number of assessments completed, and time spent on the assignments can be used to identify performance of students.

CONCLUSION

Digital systems support real-time assessment for mining information. This increases learning, transparency, and accountability, and makes it easier to evaluate trends in educational institutions. Most schools have information systems like academic performance, student discipline, attendance etc. that do not link with one another. The fragmented nature of technology constrains the integration of school information and for mining useful trends. Also, educational institutions want to format data in related ways so that results can be linked. Too often there is inconsistent terminology or coding on issues related to graduation or school dropouts. Information entered into data systems should be easy to understand and coded in comparable ways. Teachers along with parents and students will benefit from advances in research and analysis.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGEMENT

None

FINANCIAL DISCLOSURE

The authors report no financial interests or potential conflicts of interest.

REFERENCES

- [1] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Byers. Big Data.[2011] The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, May.
- [2] Felix Castro, Alfredo Vellido, Angela Nebot, and Francisco Mugica.[2007] Applying Data Mining Techniques to e-Learning Problems, *Studies in Computational Intelligence*, 62: 183-221.
- [3] US Department of Education Office of Educational Technology, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics, 2012.
- [4] Sharona Levy and IriWilensky.[2011]Mining Students' Inquiry Actions for Understanding of Complex Systems, *Computers & Education*, 56:556-573.
- [5] Joseph Beck and Jack Mostow.[2008] How Who Should Practice: Using Learning Decomposition to Evaluate the Efficacy of Different Types of Practice for Different Types of Students," Proceedings of the 9th International Conference on *Intelligent Tutoring Systems*, pp 353-362.
- [6] Vince Kellen, Cutter Consortium, Adam Recktenwald and Stephen Burr.[2013] Applying Big Data in Higher Education: A Case Study", *Cutter Consortium*, 13(8): 1-39.
- [7] Dan Sommer, Rita L. Sallam, James Richardson.[2011]Emerging technology analysis: Visualization-based data discovery tools.
- [8] McGraw-Hill.[2009] Building the Best Student Assessment Solution," New York.

**DISCLAIMER: This article is published as it is provided by author and approved by guest editor. Plagiarisms and references are not checked by IIOABJ.