Big Data in Higher Education

Prof. Deepa Jose

Abstract

Data mining and predictive analytics—collectively referred to as "big data"—are increasingly used in higher education to classify students and predict student behavior. But while the potential benefits of such techniques are significant, realizing them presents a range of ethical and social challenges. The immediate challenge con-siders the extent to which data mining's outcomes are themselves ethical with respect to both individuals and institutions. A deep challenge, not readily apparent to institutional researchers or administrators, considers the implications of uncritical understanding of the scientific basis of data mining. These challenges can be met by understanding data mining as part of a value-laden nexus of problems, models, and interventions; by protecting the contextual integrity of information flows; and by ensuring both the scientific and normative validity of data mining applications.

Introduction

The explosion of mobile networks, cloud computing and new technologies has given rise to incomprehensibly large worlds of information. The rapidly shifting dynamics of competition, coupled with a deluge of data, create new challenges for leaders across all sectors who want to tap into the power of info rmation to make better and timelier decisions [1].

Successful decision-making will increasingly be driven by analytics-generated insights. And the more accurate and timely these are the better chance a decision-maker has to anticipate and profit from change. The keys to effective data-driven decision-making are: the ability to sift through large amounts of data; and the ability to combine data from several sources to gain a more comprehensive view of the business. Seemingly unimportant data can become crucial once combined with other sources to reveal new insights [2].If we can discern the pattern in the data and make sense of what is happening, we can predict what should come next and take the appropriate action. At present, educational data mining tends to focus on developing new tools for discovering patterns in data.

Assistant Professor, YMT College of Management

In the new era of big educational data, learning analytics (LA) offer the possibility of implementing real–time assessment and feedback systems and processes at scale that are focused on improvement of learning, development of self–regulated learning skills, and student success. Visual data analytics closely involve humans to help make sense of data, from initial pattern detection and model building to sophisticated data dashboards that present data in a way that humans can act upon. K–12 schools and school districts are starting to adopt such institution-level analyses for detecting areas for instructional improvement, setting policies, and measuring results [3]. Making visible students' learning and assessment activities opens up the possibility for students to develop skills in monitoring their own learning and to see directly how their effort improves their success[4]. Teachers gain views into students' performance that help them adapt their teaching or initiate tutoring, tailored assignments, and the like[5]. Visual data analytics closely involve humans to help make sense of data, from initial pattern detection and model building to sophisticated data dashboards that present data in a way that humans can act upon. K–12 schools and school districts are starting to adopt such institution-level analyses for detecting areas for instructional improvement, setting policies, and measuring results. Making visible students' learning and assessment activities opens up the possibility for students to develop skills in monitoring their own learning and to see directly how their effort improves their success. Teachers gain views into students' performance that help them adapt their teaching or initiate tutoring, tailored assignments, and the like.

2. Challenges of using Big Data in Higher Education

2.1 Consequentialism: The Immediate Challenge

Nearly from its inception, data mining has raised ethical concerns. Once implemented, a series of challenges for both the individuals who are the subjects of data mining and the institution that bases policy on it arise as consequences. The most prominent of these are the related problems of privacy and individuality. The privacy of subjects in a data mining process is primarily a factor of information control: a subject's privacy has been violated to the extent that the opportunity for consent to collection or use of information is absent or in which personal information flows are used in ways that are incompatible with their social context. The potential of data mining to violate personal privacy spans a range of applications. Mining data allows one to infer information about the data subject that some would not be comfortable divulging themselves, and worse allows for the manipulation of or discrimination against the subject, for example, by price discrimination and restrictive marketing. These risks are very much present in higher education applications of data mining. Course recommendation or advising systems that consider student performance are a way of developing a comprehensive picture of student performance, in essence, an electronic reputation that the institution maintains and makes available to faculty and staff through dashboard and stoplight processes and administrative rules.

2.2 Scientism: The Deep Challenge

The consequential challenges of data mining are the most prominent ones, but they are not the only ones. In fact, the most difficult challenges may be ones of which institutional researchers are least aware. In the process of designing a data mining process, institutional researchers build both empirical and normative assumptions, meanings, and values into the data mining process. These choices are often obscured by a strong tendency toward scientism among data scientists. For philosophers of science and technology, the term refers (almost always critically) either to the claim that the natural sciences present both epistemologically and substantively the only legitimate way of understanding reality, or to instances of scientific claims being extended beyond the disciplinary bounds in which the claim can be supported. Such perspectives introduce the temptation to un-critically accept claims that purport to have scientific backing. Scientism has a long tradition in the social sciences, and especially in the study of education.

Scientism is a trap that, if not avoided, can do substantial harm to students. But unfortunately, current examples of data mining in higher education have embraced, rather than rejected, scientism. A non-scientistic perspective critically evaluates methods and evidence before taking action upon it. But the casual attitudes toward causality and the ignorance of even statistical uncertainty in the academic literature on data mining in higher education suggest that the authors have taken an uncritical attitude toward the underlying science of data mining. Assuming that the relationships uncovered by data mining are inherently causal and reasonably certain can lead to ineffective actions and actions that reinforce rather than interdict causal mechanisms.

3. Ethics for Ethical Data Mining

The importance of these questions, unfortunately, has not been matched by general solutions. But the above analysis suggests the formation of several fragmentary perspectives that can, if not provide solutions, lead data users in higher education to ask questions that will help refine their practices. The first step is to re-think what is meant by data mining, considering it as part of a broad technosocial system, a nexus of problem, model, and intervention built around assumptions, meanings, and values. In practice, this means thinking in terms of pol-icies in which data mining will play a role and not merely in terms of mathematical modeling techniques. The ethical questions presented in data mining will be clearer when building a data mining model is situated in relation to the perceived need for the policy, the interventions that are proposed, the expected outcomes of the policy, and the ways in which the policy will be evaluated; problems such as incompatibilities between the assumptions of the data model and those of the intervention will only be apparent from this perspective.

The empirical and normative problems presented by scientism run parallel to each other, a parallel that suggests a path toward addressing the challenge. In both cases, the question is one of whether the model's conclusion supports the interpretation given it. This is a familiar problem to empirical researchers in higher education: the problem of validity. One can thus think of scientism as an attitude that compromises (or, at the least, assumes rather than demonstrates) the validity of the problem-model-intervention nexus either empirically or normatively. Kane presents an approach to validating measures based on a series of inferences from observation to construct that can serve as a model for data mining applications.

Conclusion

There is no question that data mining can be useful in higher education. Many students' struggles with courses and programs have revealed the need for guidance that is tailored to their unique circumstances. Processes that replace general theories of how all students behave with ones that recognize their diversity while basing decisions on rigorous, reliable processes are a central tool in promoting academic success and graduation. With a wide range of social actors recognizing (for better or worse) that allowing large numbers of students to fail is an inefficient use of resources, the potential of data mining to improve the efficiency of higher education cannot be dismissed.

Data mining done well presents challenges to both individuals and institutions, and because of scientistic attitudes it is often done poorly at great cost, both practically and morally. Institutional researchers must minimize this risk. To do so, institutional researchers must understand data mining as part of a technosocial whole that spans the entire policy process. They must ensure the contextual integrity of information flows to protect the actors involved in data mining. And they must ensure both the scientific and the normative validity of the data mining process. Done properly, institutional research can secure these gains without com-promising its commitment to the good of students.

References

- *Ayesha, Shaeela, T Mustafa, AR Sattar, and MI Khan. "Data Mining Model for Higher Education System." Eu-ropen Journal of Scientific Research 43, no. 1 (2010): 24–29.*
- *Baepler, Paul, and Cynthia James Murdoch. "Academic Analytics and Data Mining in Higher Education." Inter-national Journal for the Scholarship of Teaching and Learning 4, no. 2 (2010). http://academics.georgi-asouthern.edu/ijsotl/v4n2/essays_about_sotl/PDFs/_BaeplerMurdoch.pdf.*
- *Baez, Benjamin, and Deron Boyles. The Politics of Inquiry: Education Research and the "Culture of Science." Albany, NY: State University of New York Press, 2009.*