Data Analytics and AI in the Education and Online Learning Landscape

Ryan Spellman

College of Communication and Information, Kent State University

LIS-60010: The Information Landscape

Dr. Sofia Empel

November 14th, 2020

The Information Landscape in the Neighborhood of Education and Online Learning Where Big Data and AI Have Moved In

The information landscape in education and online learning is a very sensitive place right now. In the Academic College and Research LIbraries' (ACRL) 2020 top trends, they mention that librarians are beginning to question the usefulness of learning analytics, especially when weighing that against the effects such data retention has on student privacy (Research Planning and Review Committee, 2020). Kamenetz (2014) brings up that students are free to leave social network sites if they are unhappy with privacy policies, but are often forced to use educational resources that collect their data. The amount of necessary online learning tools is also only increasing due to the global pandemic environment, so this concern is very poignant. On top of that, privacy concerns are not the only contentious issue with big data and AI in education.

Looking at Course Signals, we see an example with good intentions that can cause harm. Course Signals is a program that takes a student's accumulated points, the time they spend on tasks, and their past performance in a particular course, translating it into a red light for high risk of failing, yellow light for medium risk, and green for low risk, which is then displayed on Blackboard for the student and also available for instructors to review (purdueidc, 2011). This seems like it would be helpful, but consider some other implications.

Kamenetz (2014) points out two factors for consideration on how Course Signals might be harmful for student success: stereotype threat and the Pygmalion Effect. The stereotype threat refers to research showing that stereotyped minority test takers underperform (Steele, 1995). Here, Kamenetz is raising the question of whether seeing

a red flag will increase a student's rate of underperformance or worse, failure. The Pygmalion Effect refers to research that shows an instructor's positive expectations of a student affects their outcome (Caffrey, 2020). If a student starts off a little bumpy, and another starts in the green, will this bias the instructor's treatment of one versus the other?

Key Data Analytics Technologies in the Education Landscape

The education landscape is vast and has technology deeply embedded into many of its facets. This makes it easy to see where three of the four types of analytics technologies covered by Guo (2017) in a discussion on machine learning are key: descriptive, diagnostic, and predictive analytics. For the sake of brevity, we will focus on one common Learning Management Software (LMS) solution for evidence of this fact: Blackboard Learn.

Blackboard Learn is one of four dominant LMS systems in the United States and Canada (Hill, 2018). As such, it is expected that they will have robust offerings such as big data analytics. Most of our evidence of what data analytics technologies are employed by them can be found in their Blackboard Learn Analytics suite.

Descriptive analytics involve what happened in the past, and are typically displayed in charts, graphs, and other human-interpretable visuals (Guo, 2017). For example, we graph door count data to see if our hours are meeting the needs of our students where I work. In our Blackboard example, the Activity Matrix and Activity Grade Scatter Plot provided Dr. Selim, a UAE professor, the opportunity to review performance data and easily find at-risk students (Blackboard, 2019).

Diagnostic analytics is just as it sounds, analytics processes that seek to

understand the why of the data (Guo, 2017). One of the top questions that Blackboard Analytics for Learn purports to address is, “What student activities are correlated with desired outcomes like grades and course completion?” (Blackboard, 2018). This is a prime example of the use of diagnostic analytics in the education landscape to automate correlations.

Predictive analytics seeks patterns to make predictions (Guo, 2017). Blackboard Learn also offers predictive analytics. Blackboard Engage (formerly Blackboard Predict) is a package that predicts which students are, or will be, at risk by incorporating data from the institution’s SIS and LMS into the analytics process (Blackboard, n.d.).

Changes in Other Information Intensive Areas of Our Society

Beyond the Education Landscape

Something that piqued my interest was the discussion about how General Motors has a patent on billboards that will change based on data scraped from the vehicle passing by, tailoring it based on factors such as where they are going, how long they have been on the road, and possibly even age and gender via voice analysis (The Checkout, 2013). This got me thinking on how gas stations are beginning to install little flat screen displays at the pump so they can bombard customers with adverts.

If you combine these screens with the data such locations are mining from their rewards points system, which are more often than not swiped at the pump, the likelihood of targeted adverts is incredibly high. Then, what if data from credit/debit card purchases or other rewards systems are integrated together? Through the use of diagnostic and predictive analytics, these systems could infer from purchasing patterns that at this time of day you usually eat dinner, and that you love stopping by Subway,

and therefore hit you with a Subway advert or even hit you with a coupon for further encouragement.

As noted by Drysdale, stores have been analysing consumer purchase patterns through rewards programs and custom tailoring vouchers for them to encourage patronage (The Checkout, 2013). I feel this is certain to continue to escalate, and perhaps has without our knowledge as we rashly agree an accept the many terms of agreement that are pressed upon us when we install apps, sign up for accounts, and join rewards programs. I feel the data deluge of our consumer activity is going to continue to see change and raise many issues as predictive, and even prescriptive, analytics continues to improve and be used to influence consumer behavior patterns.

Brief Narrative of the Ethical Questions Raised by These Technologies and How the Information Profession May Play a Role in the Future

In the information profession, and particularly in libraries, privacy is a hallmark of our service. Privacy is the main ethical concern that I see on the surface here. The American Library Association’s (ALA) Bill of Rights, Article VII, asserts that “all people, regardless of origin, age, background, or views, possess a right to privacy and confidentiality in their library use" (American LIbrary Association, 2008). Data analytics is exciting and can do amazing things, but we must be cautious stewards of its use in librarianship. How do we use it to improve services for our communities, yet not overstep into privacy invasion? And what of broader social contexts, do privacy education concerns fall into the purview of information literacy, something that has long been established as a responsibility of librarians?

In reference to data analytics that help libraries make operating decisions, the

ALA has a rather robust FAQ item that provides advice on how libraries can use data analytics to help improve services for their patrons, but still safeguard privacy. As previously mentioned, ACRL also listed student privacy concerns over data analytics in their 2020 top trends (Research Planning and Review Committee, 2020). It is clear that this is not an unfounded concern and will need attention moving forward. But how do we do so?

ALA provide four key points to consider that will help safeguard patron privacy and confidentiality:

- only collect data that is needed,
- do not collect “high risk” data such as immigration status,
- ensure vendors follow best practices and establish addendums they must agree to,
- alter data through aggregation or obfuscation to protect patron identity (American Library Association, 2007).

Working in libraries means we should use every advantage to make our services better for the user, but we must also protect our users. For example, collecting data on how a user browses the catalog could easily be anonymized or even collected in an outright anonymous way. This could lead to major insights, from collection development considerations to user interface improvements. Using ALA’s suggestions, especially as data aggregation and analytics improves, we must ensure that the platform being employed follows best practices, that we understand what data they are storing, and also go over the fine print on what they are doing with the user data.

My second question relates to contexts outside of the library and what

responsibilities we have in educating our communities on data and privacy concerns. For me, personal privacy in this data driven landscape is part and parcel of information literacy. We make information literacy efforts so our communities can learn how to evaluate sources and avoid harmful disinformation. Similarly, it only makes sense that we make an effort to educate our communities on what is happening behind the scenes and how they can protect their privacy.

ACRL has a privacy workshop as part of its Information Literacy Sandbox repository that focuses on educating new college students about how to develop digital behaviors that align with the core library values of privacy and intellectual freedom (Chisholm & Hartman, 2018). The workshop framework covers everything from software update security to reading terms of service in plain English. With data analytics becoming such a powerful tool and the increase in machine learning and AI sophistication, it is integral our communities are equipped with best practices to safeguard their privacy. This involves not only educational outreach as seen in the ACRL framework example, but also continuing education on the part of library professionals. In order to educate our communities on what happens to their data and how they can protect it, librarians must be up for the challenge of staying abreast on current trends and practices in the data analytics field.

References

American Library Association. (2007, May 29). *Privacy and confidentiality Q&A.* http://www.ala.org/advocacy/intfreedom/privacyconfidentialityqa

American Library Association. (2008, June 13). *Privacy.* http://www.ala.org/advocacy/privacy

Chisholm, A. & Hartman-Caverly, S. (2018). *Privacy workshop.* ACRL Framework for Information Literacy Sandbox: A Platform and Repository for Sharing Framework Materials. https://sandbox.acrl.org/library-collection/privacy-workshop

Blackboard. (n.d.). *Blackboard engage: Student predictive analytics.* https://www.blackboard.com/teaching-learning/data-and-analytics/blackboard-engage

Blackboard. (2018). *Blackboard analytics for learn: Digital learning environments generate mountains of data.* https://www.blackboard.com/resource-node/blackboard-analytics-learn-digital-learning-environments-generate-mountains-data

Blackboard. (2019). *How 7 higher education institutions in the MEA region use analytics to improve institutional performance and student success.* https://www.blackboard.com/en-mea/resources/case-study-a4l-mea

Caffrey, C. (2020). Pygmalion effect (Rosenthal effect). *Salem Press Encyclopedia*.

Checkout. (2013, April 11). *Data mining* [Video]. Youtube. https://youtu.be/f2Kji24833Y

Guo, Y. (2017, August 24). *What is machine learning?* [Video]. Youtube. https://youtu.be/HcqpanDadyQ

Hill, P. (2019, February 5). *State of higher ed LMS market for US and Canada: 2018 year-end edition.* eLiterate. https://eliterate.us/state-higher-ed-lms-market-us-canada-end-2018/

Kamentez, A. (2014, July 4). *Big data comes to college.* NPR. https://www.npr.org/sections/ed/2014/07/04/327745863/big-data-comes-to-college

purdueidc. (2011, October 9). *Course signals explanation* [Video]. Youtube. https://youtu.be/KxYewO9iAgw

Research Planning and Review Committee, A. (2020). 2020 top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News, 81*(6), 270. doi:https://doi.org/10.5860/crln.81.6.270