Old Dominion University
ODU Digital Commons

STEMPS Faculty Publications | STEM Education & Professional Studies

2019

The Role of Data Analytics in Education: Possibilities & Limitations

Robert L. Moore
Old Dominion University, robmoorephd@gmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/stemps_fac_pubs

Part of the Artificial Intelligence and Robotics Commons, Educational Technology Commons, Elementary Education Commons, and the Higher Education Commons

Original Publication Citation
Moore, R. (2019). The role of data analytics in education: Possibilities & limitations. In B. H. Khan, J. R. Corbeil, & M. E. Corbeil (Eds.), *Responsible Analytics and Data Mining in Education: Global Perspectives on Quality, Support, and Decision Making* (pp. 101-118). New York: Routledge, Taylor & Francis Group.

This Book Chapter is brought to you for free and open access by the STEM Education & Professional Studies at ODU Digital Commons. It has been accepted for inclusion in STEMPS Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

The role of data analytics in education: Possibilities and limitations

Book Chapter in: *Responsible Analytics and Data Mining in Education: Global Perspectives on Quality, Support and Decision Making.*

Robert L. Moore – robmoorephd@gmail.com | Twitter: http://www.twitter.com/robmoore3 | ORCID: http://orcid.org/0000-0002-5645-9297

Citation:

Moore, R.L. (2019). The role of data analytics in education: Possibilities and limitations. In B. Khan, R. Corbeil, & M. Corbeil (Eds.), *Responsible Analytics and Data Mining in Education: Global Perspectives on Quality, Support, and Decision-Making* (pp. 101-118). Routledge, New York. DOI: http://doi.org/10.4324/9780203728703-8

Comments:

This is an original manuscript/preprint of an article published by Taylor & Francis in *Responsible Analytics and Data Mining in Education: Global Perspectives on Quality, Support, and Decision Making* in 2019, available online: https://www.routledge.com/Responsible-Analytics-and-Data-Mining-in-Education-Global-Perspectives/Khan-Corbeil-Corbeil/p/book/9781138305908

Abstract: In the last decade, we have seen dramatic increases in the integration of technology within education. It has now become commonplace for K-5 educators to apply learning management systems (LMS) in ways that were previously only seen in higher education contexts. Similarly, on the higher education side, we are seeing a significant increase in online learning evidenced by the growing number of for-profit online colleges and universities (Picciano, 2012). This chapter utilizes Khan's Learning Framework (Khan, 2001, 2005) to explore the role data analytics can play in education by looking at the possibilities and limitations of analytics.

The Role of Data Analytics in Education: Possibilities and Limitations

Robert L. Moore

Orcid.org/0000-0002-5645-9297

Abstract

In the last decade, we have seen dramatic increases in the integration of technology within education. It has now become commonplace for K-5 educators to apply learning management systems (LMS) in ways that were previously only seen in higher education contexts. Similarly, on the higher education side, we are seeing a significant increase in online learning evidenced by the growing number of for-profit online colleges and universities (Picciano, 2012). This chapter utilizes Khan's Learning Framework (Khan, 2001, 2005) to explore the role data analytics can play in education by looking at the possibilities and limitations of analytics.

Introduction

In the last decade, we have seen dramatic increases in the integration of technology within education. It has now become commonplace for K-5 educators to apply learning management systems (LMS) in ways that were previously only seen in higher education contexts. On the higher education side, we are seeing a significant increase in online learning evidenced by the number of for-profit online colleges and universities (Picciano, 2012). While technology can be described as a disruptor, it can also be seen as the solution to solving the new challenges it presents. (Picciano, 2012). For instance, one of the benefits of integrating an LMS is the ability to capture interactions and activities of students in new ways. This variety of data, sometimes referred to as 'trace data,' includes data points such as the number of logins, what time of the day the logins occur, and which resources are used. With the advent of this data collection, there is now a calling for instructors to successfully interpret this data and understand how it can inform their instruction. With the use of this tool, institutions are also generating prodigious amounts of student data (Asif, Merceron, Ali, & Haider, 2017; Daniel, 2015; Picciano, 2012; Roberts, Howell, Seaman, & Gibson, 2016). This data generation is coupled with advances in the database technology that can allow for both storage and analysis of these data points (Daniel, 2015). But with so much data, the largely unanswered questions of what data is being collected, how it is being used, and who should have access to the data still remain.

With this newly accessible data, educational institutions can now leverage this information to make decisions – both administratively and pedagogically (Daniel, 2015). This type of data-based decision making is not new. As Picciano (2012) emphasizes, it has been around since the 1980s, but what is new are our technological advances and how we can now handle the vast amount of data that these learning management systems have created. The cost and portability of storage has significantly impacted how we view and manage our data. In the early 2000s, having an external drive that could hold 750MB was an add-on for many personal laptops; now basic laptops have

250GB of internal hard drive space. This more complex analysis is often referred to as Analytics (Picciano, 2012). Institutions can take data from multiple locations (e.g. learning management systems or customer relationship management systems) and analyze this data in an effort to address student retention and improve support services (de Freitas et al., 2015). It is for this reason that we see data being viewed as a type of institutional currency which can be a valuable tool in the recruitment, retention and eventual matriculation of students (de Freitas et al., 2015).

Khan's Framework

Institutions need to be strategic not only in their collection of data, but in their use of data. It is vital that they consider their goals for the use of the data and tailor their systems to meet these goals. Just as there is a plethora of ways to use, collect and analyze data, there are an equally large number of ways to frame the analysis and discussion. For the purposes of this chapter, the lens of data analytics, specifically the role that it can play in education, will be through Khan's Learning Framework (Khan, 2001, 2005). This framework will discuss the role that data analytics can play in the decision-making processes within all educational contexts, with a specific focus on higher education institutions. This framework features eight dimensions that delve into issues that organizations should consider as they implement a new initiative; in this case the use of data analytics. These dimensions are pedagogical, technological, interface design, evaluation, management, resource support, ethical and institutional. For a more detailed discussion of all eight dimensions, see Chapter 2 of this volume. This chapter will focus specifically on four of the dimensions – pedagogical, technological, interface design, and ethical and are highlighted in Figure 1.

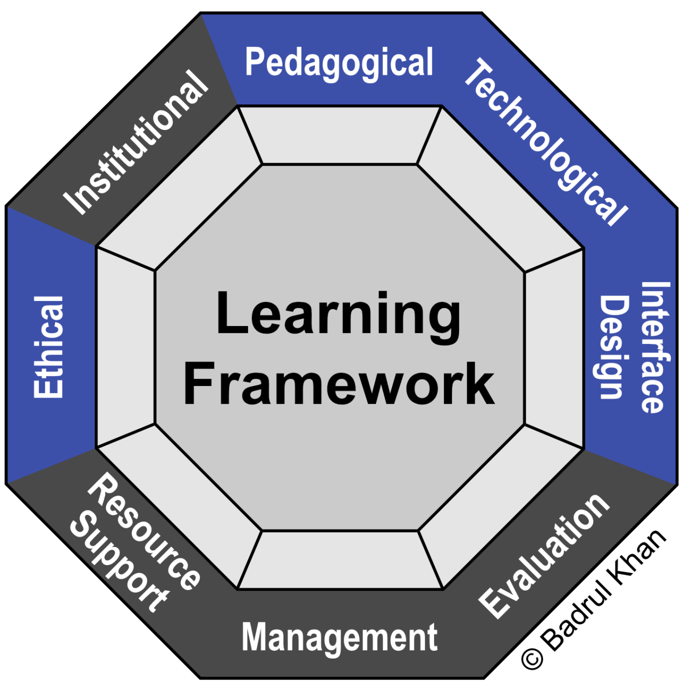

Figure 1. Eight dimensions with highlights on chapter dimensions. Adapted from Khan (2001, 2005).

Pedagogical Dimension

The first of the four dimensions discussed in this chapter is the pedagogical dimension. In Khan's framework (Khan, 2001), this dimension is chiefly concerned with the learners' needs and how the learning objectives can be reached through the effective design, delivery and implementation of the instructional content. The learning environment is a key component for this dimension and it is important to ensure the environment is facilitating the necessary knowledge transfer effectively. This dimension has relevance to the chapter's focus along the lines of the role of data analytics and data mining, and how educators and administrators can make use of these analyses to inform decisions. To explore these concepts, the next section looks at learning analytics, educational data mining, texting mining and the associated implications and uses within educational contexts.

Papamitsiou and Economides (2014) provide a useful distinction between learning analytics and educational data mining. In their taxonomy, learning analytics would be defined as the 10,000-foot view of analysis; this contrasts with educational data mining which is looking for patterns at the individual level and from there looking to make decisions. They elaborate in explaining that learning analytics and EDM provide the ways to gather, process and make decisions on data in a way that can advance the educational environment (Papamitsiou & Economides, 2014). Branching from EDM is Text mining, a specific type of data mining focused on text. This type of mining is particularly relevant because of the learning management systems and the prevalence of online discussion boards in online learning (Gupta & Lehal, 2009; Martin & Ndoye, 2016; Picciano, 2012) which generate tremendous amounts of text-based data.

Learning Analytics

The literature contains multiple definitions of learning analytics, but Siemens (2013)'s definition has the most relevance for this discussion. In his definition, he makes the connection between the collection of the different student data points and the use of that information to inform decisions regarding student retention and learning behaviors. Additionally, Siemens' definition looks at the way learning analytics can be used to provide feedback that is personalized for the students. It is at this personalization level that learning analytics can have a tremendous impact on students' experience. One of the common ways we see the use of learning analytics is through monitoring a student's participation and performance within a course (Picciano, 2012). This may be one of the reasons that (de Freitas et al., 2015) identifies learning analytics as an emerging area of research within the educational science community and as pointed out by Arnold and Pistilli (2012), the effective leveraging of learning analytics is one way that institutions can make sense of the large volumes of data being collected.

The practical applications and utility of learning analytics are numerous and can benefit both online course instructors and students (Martin & Ndoye, 2016). As mentioned, a common use is to monitor a student's performance, but the benefit of learning analytics comes from *how* that information can be used. And as Arnold and Pistilli (2012) share, instructors can take this information and use it to provide timely and relevant feedback to students. However, while learning analytics in general can be very helpful to an instructor in supporting student learning, Martin and Ndoye (2016) caution that instructors should be cognizant of the fact that there are limitations to the effectiveness of learning analytics on non-online contexts. In the K-12 context, learning analytics has been used for teacher evaluation (e.g. Grossman, Cohen, Ronfeldt, & Brown, 2014) and to explore and predict teacher attrition (e.g. Henry, Bastian, & Fortner, 2011). In contrast, learning analytics is used in higher education to focus on students – from their retention to their learning behaviors and as a tool to provide a level of personal support and feedback from their instructors (Roberts et al., 2016).

Educational Data Mining

While learning analytics is focused on the system-wide analysis, educational data mining, or EDM, looks at the individual as the unit of analysis. Within analytics work in education, data mining is where a significant amount of work and research is being conducted (Daniel, 2015). Berland, Baker, and Blikstein (2014) further explain the utility of data mining, specifically educational data mining (EDM), in being able to support research and inform policy makers. The term data mining is commonly used to refer to any process that results in searching or "digging into" a data file for information to better understand a particular phenomenon (Picciano, 2012, p.12). It is not a new concept to mine for information in data, but using these approaches in educational contexts is an emerging area of research (Baker & Yacef, 2009; Kabakchieva, 2013). There are various techniques of data mining, such as classification and clustering, which when applied correctly allow institutions to extract meaning and significance from student data (Asif et al., 2017). These techniques are discussed by (Baker, 2010), who presents five primary categories or approaches to EDM: prediction, clustering, relationship mining, discovery within models, and visualization. Of these categories, Asif et al. believe that the application area of prediction in EDM is anticipating student educational outcomes. Baker (2013) proposes predictive models in educational data mining context are intended to infer a single aspect of the data (the predicted variable, akin to dependent variables in traditional statistical analysis) from some combination of other aspects of the data (predictor variables, akin to independent variables in traditional statistical analysis).

Text Mining

Text mining differs from educational data mining because text "can work with unstructured or semi-structured data sets such as emails, full-text documents and HTML files" (Gupta & Lehal, 2009, p. 60). The ability to analyze written text is incredibly important with the increasing reliance on learning management systems and online discussion forums. Gupta and Lehal (2009) define text mining as "the process of extracting interesting and non-trivial information and knowledge from unstructured text" (p. 60). But text mining is more than just superficially looking at the data; it is about making sense of it and using it to inform instructional decisions. These decisions can range from identifying areas where students are struggling to offering remediation steps or attaining a better understanding of what resources are most valuable (and conversely not as valuable) to students to inform instructional design considerations (Papamitsiou & Economides, 2014).

One of the benefits text mining can offer is the ability to understand what previously went unseen by an instructor. Through the analysis, the instructor can offer remediation to students that may not even realize that they need help. Thus, it fosters a way to support the adaptive and iterative nature of good instruction by providing formative evaluation and assessment that can then inform future instructional decision that can make instruction more efficient and engaging for students in all instructional contexts. Text mining offers a systematic way to analyze and understand a large corpus of text. As more data is collected through LMS or other mediums, such as social media, it becomes incredibly useful to have an effective way to synthesize this information. Gupta and Lehal (2009) identify several means of text mining, including, "information extraction, topic tracking, summarization, categorization, clustering, concept linkage, information visualization, and question answering" (p. 61). While this classification is framed by Gupta and Lehal as being specifically for text mining, these categorizes can also be applied to the overall discussion and understanding of the use of educational data mining and learning analytics.

Information extraction. With information extraction, software is leveraged to look for relationships within the text which make it an ideal choice when dealing with large volumes of text (Gupta & Lehal, 2009). This technique is particularly useful when the corpus is not already structured, e.g.emails or websites, and relationships need to be established before moving forward with the analysis. This technique can help make the design decisions for more in-depth analyses of the data being researched.

Topic tracking. For topic tracking, users need the ability to create a profile that can monitor user interests in topics and predict other documents that match these interests (Gupta & Lehal, 2009) An example of topic tracking is Google alerts, where one can select to receive alerts on certain search terms or phrases. These alerts provide a useful way to stay updated with new information. However, as Gupta and Lehal caution, it can also result in unnecessary information depending on the structure of search terms. Still, these alerts will help ensure that individuals are receiving the latest references as it can signal when new work has been contributed to the field.

Summarization. Another useful technique for culling large amounts of data is summarization, which can narrow down the data to only the parts that meet certain criteria. From this shorter list, the researcher can work on researching the topics that are of the most interest. This technique is useful when the researcher is interested in first determining the viability and usability of the corpus for specific research goals. Gupta and Lehal (2009) explain that, in the time that it would take a human to read just one paragraph of text, a software program can analyze and summarize a book's worth of text. Consider an instructor that is interested in finding out what their students have contributed in the discussion forums. Having the ability to quickly summarize the information from over thirty students' posts can aid the instructor in making real-time adjustments to their course.

Categorization. For a researcher using a deductive research approach, categorization can be useful as it "involves identifying the main themes of a document by placing the document into a pre-defined set of topics" (Gupta & Lehal, 2009, p. 63). An example of categorization is the means through which many websites create their knowledge bases for help documents. An end-user can select a category that best characterizes their problem and sort through information more efficiently. As more help documents are added to the system, they are categorized based on previous documents, creating new relationships and linkages that are then sortable by the end user.

Clustering. Clustering is a variation of categorization. The key difference between categorization and clustering is that clustering is accomplished automatically in document repositories. One is more likely to see clustering when there is a large volume of documents, such as usage reports for an online course.

Concept linkage. Gupta and Lehal (2009) explain that concept linkage is especially useful when there are large amounts of data and it is not plausible for a researcher to know every linkage or connection, e.g. in the biomedical fields. Being able to quickly make linkages to past tests or clinical trials could ultimately mean the difference between life and death for a medical patient.

Information visualization. Information visualization provides a way to visually depict the information that has been analyzed. Through this visualization, new connections and relationships may emerge. Brown, Cowan, and Green (2016) documented how they used text mining techniques to link social media and its impact on faculty productivity. They looked at one faculty member’s social media impact and connections and categorized the connections by different social media networks, presenting their findings in a visual depiction (Figure 2).

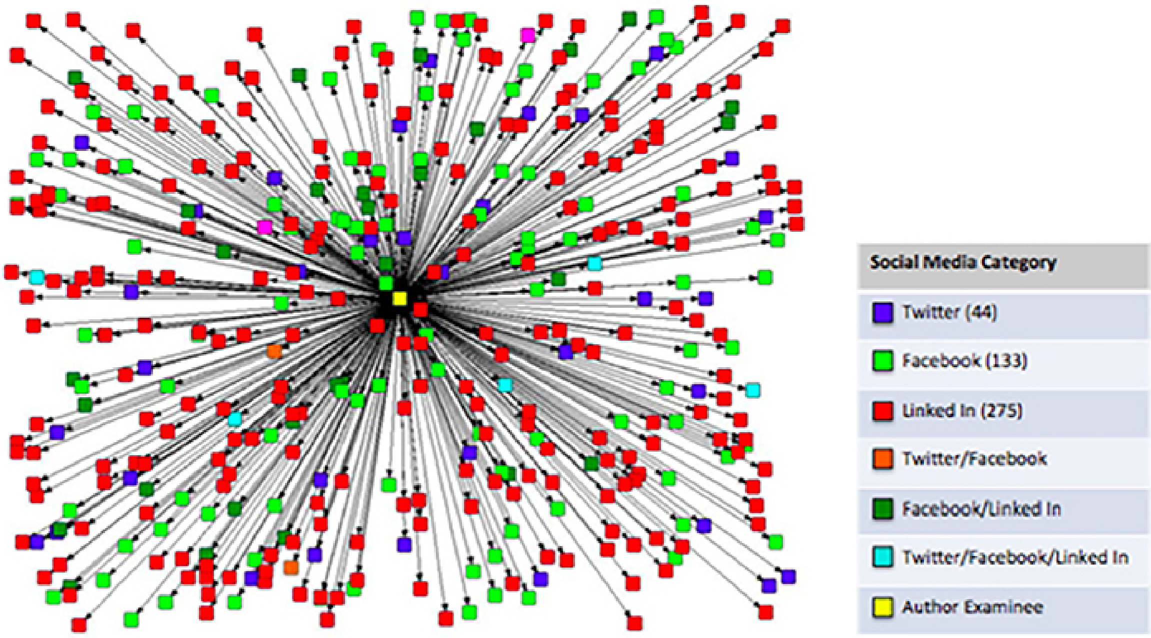

Figure 2. Author examinee's composite social media network. Source: Brown, Cowan, & Green, 2016.

Question answering. This technique is demonstrated in many websites' frequently asked questions or knowledge base sections of their site. In these areas, a visitor types in a question or the start of a question and the site returns results that match the query. On the back end, the website tracks what searches occur and what search terms are being used. The development team can access these logs and can modify and refine their tags or taxonomies that will improve the search experience for the end user.

Data-Driven Decision Making

Each of these techniques – whether it is learning analytics, EDM or text mining – is creating the opportunity for an institution to be able to make a decision. And the use of data to make a decision is what is defined as "data-driven decision making" (Picciano, 2012, p. 11). In the higher education context, student retention is where we see the most application of data-driven decision making (Picciano, 2012). It can also be used to facilitate student learning (Martin & Ndoye, 2016; Roberts et al., 2016). On the K-12 side, as previously noted, we are seeing the application typically with either teacher evaluation or to explain teacher attrition (Grossman et al., 2014; Henry et al., 2011).

Technological Dimension

In the second dimension, the focus shifts from the content to the technological requirements that include the learning environment and the tools necessary to deliver it. This dimension also addresses hardware and software requirements, as well as infrastructure planning. Technical requirements such as server capacity, bandwidth, security, backups, and other infrastructure issues are also addressed. This is a key consideration regarding data mining and analytics, as the technological dimension also addresses networking infrastructure issues relating to data volume and transmission.

Big Data

A discussion about data analytics cannot happen without a discussion of the buzzword, ‘big data’. Similar to learning analytics, there are multiple definitions of this term. Picciano (2012) describes big data as a generic term that assumes that the information or database system(s) used as the main storage facility is capable of storing large quantities of data longitudinally and down to a very specific transaction. Big data is not just describing the size and mobility of the data, but referring to the techniques and technologies used to analyze those large datasets (Daniel, 2015; Yang, 2013). As Yang (2013) points out, the analysis of huge amounts data is not new, but we now have methods and technologies that can analyze these volumes of data at never-seen-before speeds. It is this ability to take data – both structured and unstructured – and turn it into a format that is useful and has a value that this chapter will refer to as big data (Daniel, 2015). In a seminal work regarding big data, Daniel (2015) identifies the fundamental characteristics of big data as: volume, velocity, veracity, variety, verification, and value (p. 908).

Big data has significant technological impacts for educational institutions. Big Data can influence higher education practice, from enhancing students experiences and improved academic programming, to more effective evidence-based decision making, and to strategic response to changing global trends (Daniel, 2015). These impacts can be understood through the three stages identified by Daniel as data collection, data analysis and visualization and application. As previously mentioned, institutions are using more online tools which has made the data collection easier. With these data points, institutions can make use of predictive tools that can improve learning outcomes for individual students (Daniel, 2015). Coupling predictive analytics with big data provides the ability to identify *useful* data and turn it into useful information by identifying patterns and deviations from patterns (Daniel, 2015).

Big data and analytics can become part of the solutions and be integrated into administrative and instructional functions (Picciano, 2012). Big data concepts and analytics can be applied to a variety of higher education administrative and instructional applications, including recruitment and admissions processing, financial planning, donor tracking and student performance monitoring (Picciano, 2012). Big data can be used for learning analytics purposes at a range of levels within the university including: university wide models predicting retention (de Freitas et al., 2015) and course level data providing feedback on learning of a particular subject to individual students (Arnold & Pistilli, 2012; Daniel, 2015)

But while there are many benefits of using big data, the application of big data to learning analytics for the purposes of learning instruction is less common (Dede, Ho, & Mitros, 2016). (Picciano, 2012) identifies three key concerns regarding the use of big data and learning analytics. The first concern is that the most effective application of learning analytics is with online courses since that allows for more data points to inform decision making. A second concern is that there needs to be an availability of people with the proper training and resources in order to accurately and effectively analyze the data. Finally, due to privacy being a major concern, precautions need to be taken to ensure that the extensive data collections of student instructional transactions are not abused in ways that potentially hurt individuals.

Machine Learning

Machine learning is a method through which big data can be analyzed and turned into the actionable information that can be of benefit to an organization (Daniel, 2015). A challenge a computer may face is that it does not know the nuances of language, and may thus fail to accomplish the same connections a human would easily piece together. One way to mitigate this is through the use of algorithms in which text mining is explained through the use of machine learning. These algorithms can be leveraged in either a supervised or unsupervised approach. For supervised machine learning, we have an idea of our model and attempt to feed data into the model to make predictions. With unsupervised machine learning, we do not have our model in mind and instead, task the computer with organizing the data into groups or clusters, and form our interpretations from these stratifications.

Supervised vs. unsupervised. Two examples help illustrate the differences between these two approaches. The first example is how email systems, such as Outlook or Gmail, handle spam. These are accomplished with a supervised machine learning approach as there are certain criteria for what is considered spam which is usually based on previously established examples of spam. However, it is machine learning because the user is able to ‘train' their email system to better identify spam. One can mark an email as not being spam, or whitelist a particular sender. Over time, the email system will learn how to make a better prediction and improve the model. Another example is Netflix or Amazon. Both of these companies have been able to track individual user preferences such as: what movies have been watched, what items have been highly rated and purchased in comparison to other users, and have used this information to make recommendations on what the next movie or purchase should be. As the user’s interactions with the sites change, so too does the algorithm and the resulting recommendations. An example of unsupervised machine learning was outlined by Ezen-Can, Boyer, Kellogg, and Booth (2015) in their study of discussion forums in an MOOC-Ed course. For their study, they had a large volume of text and did not use an *a priori* coding approach, but instead wanted to extract the meaning from the text and had their algorithm predict and organize the content into clusters. From there, they analyzed the clusters and derived their meaning of the content within these forums.

Discussion Forums. Discussion forums are typically the most active place within an LMS and have the least structured data and text that are created, particularly in an online course. Ezen-Can et al. (2015) point out that there has been a marked increase in the interest in using learning analytics to understand better student activities within

MOOCs. Specifically, they state, "one very important source of data in MOOCS is the textual dialogue among students…on discussion forums" (p. 146). For an online course, there may be weekly assignments to post original thoughts and replies to the discussion forum. These forums are also where a significant amount of the student-student interaction takes place, and much of the learning in an online course can be found in these spaces, particularly in MOOCs (Wong, Pursel, Divinsky, & Jansen, 2015). Using an unsupervised machine learning approach can determine which forums are most popular, what types of information is being created and shared, and what is resonating the most with students based on their engagement with particular topics, forums, and students.

Face-to-Face Contexts. There is also potential for machine learning for face-to-face learning environments. Common data collection points for a course are teaching evaluations, which are typically summative in nature. By leveraging SMS texting, Leong, Lee, and Mak (2012) present an example of how SMS texting can allow students to give more immediate and formative evaluations of their instructor and propose the use of a sentiment analysis to parse these evaluations and derive meaning. Additionally, while this more timely feedback is useful, it can be difficult to "extract insights from an analysis of such noisy and unstructured data in SMS texts" (Leong, Lee, & Mak, 2012, p. 2584). To mitigate this, Leong et al utilized an exploratory data analysis at the concept level which “involves viewing a list of concepts extracted, the relevant statistics in terms of frequency and percentage of occurrence of the respective concepts as well as the number and percentage of documents in the corpus that contain the concepts” (p. 2585). Their research illustrates the potential for additional future research in this area. A particular limitation of their study is the amount of pre-analysis data cleanup that needed to be done to clean the data, and the tendency of texts being incomplete in nature and subsequently harder to analyze. However, they did find that timely information is helpful for the instructor. It would be interesting to remake their study while using a web-based tool such as PollEverywhere or Qualtrics to perform these timely evaluations and examine if the limitations of SMS text (e.g. emoticons, spelling errors, and limitations in the amount of text, etc) would factor into the results as much as in their current study.

Another example of using machine learning within a face-to-face classroom setting is outlined in a study conducted by Ai, Sionti, Wang, and Rosé (2010) on transactivity. They explain that transactivity is “well studied in the domain of educational psychology” and elaborate “[that] transactive contributions are arguments constructed in such a way as to reference... the previously expressed reasoning of self or others” (p. 976). In their study, they transcribed 185 minutes of a middle school math classroom. Their research is based on “the assumption that an assessment of level of trans-activity in an in-progress discussion would be valuable information for facilitators to use in deciding what is needed from them to help keep the conversation on a productive path” (Ai et al, 2010, p. 977). They leveraged three supervised machine learning algorithms to help organize the transcriptions of the class, specifically “Naïve Bayes, support vector machines, and decision trees” (p. 980). Their research provides some new opportunities for both students and instructors, particularly in face-to-face learning environments. For students, this trans-activity research could be used to track their progress over time, particularly in the areas of critical thinking and debate. For instructors, it can highlight the best ways to fully engage students in classroom discussions and help move the classroom to a more student-centered and collaborative learning environment. If instructors better understand how to facilitate class discussions, they can make the learning environment more inclusive and engaging for students – and this research can help to move it in that direction. The main limitation for this is the need to transcribe classroom activities, and developing code for different instances to build the connections and leverage the supervised algorithms.

Interface Design

The third dimension, interface design, is concerned with factors related to maximizing usability and the user experience. Factors such as web design, content design, navigation, accessibility, and usability testing are addressed in this dimension. The interface design dimension also addresses accessibility and usability issues pertaining to data portals by helping decision-makers address making the design accessible for stakeholders. As highlighted below, one of the ways that this is done is through the use of dashboards.

Student Involvement

As this dimension is interested in the design and usability of the system, it is critical that stakeholders, in this case students, are involved. It is of utmost importance that students are engaged through the process early enough in order to ensure that the systems are meeting their learning needs (Gašević, Dawson, & Siemens, 2015; Kruse & Pongsajapan, 2012; Slade & Prinsloo, 2013). But while it is well documented that students need to be involved in the decision-making process (see Slade & Prinsloo (2013), it is less clear exactly how and when to make this happen. Daniel (2015) suggests the speed with which technology has evolved and the demands on higher education institutions to develop and deliver high-quality programs has outpaced the ability to adequately involve and engage students in the decision-making processes. And while this may be true, (Greller & Drachsler, 2012) warn that basing judgements on a limited set of parameters has the potential to create a context for profiling, which can result in limiting students' potential and damaging self-efficacy.

Dashboards

One way that data analytics can be used to support student learning is using dashboards. These dashboards provide students with timely, or depending on the system, real-time feedback. Pardo and Siemens (2014) contend that there are significant benefits for students when they can have real-time feedback. They continue to explain that by giving students more timely feedback, it better prepares students to take corrective action and can therefore result in higher achievement levels for the students. Additionally, Roberts et al. (2016) found that dashboards allowed students to engage in self-regulated learning.

The use of dashboards also supports a concept called “academic analytics” which specifically looks at how to identify and support underperforming students (Arnold & Pistilli, 2012). The hope of academic analytics is that students are identified early enough for corrective action to be made. To explore the utility of academic analytics, Arnold and Pistilli (2012) investigated the use of a tool called Course Signals at Purdue University. One of the core utilities of Course Signals was its use of predictive modeling to allow faculty to contact students before the students got too far behind (Arnold & Pistilli, 2012). This tool allowed faculty to send messages to students regarding their current academic performance in a specific course. But it also allowed for faculty to direct students to available resources that can help them remediate issues that they were observing. There were several observed benefits of Course Signals. For one, courses that used Course Signals had higher grades and fewer withdrawals (Arnold & Pistilli, 2012). Also, students felt that Course Signals were a helpful and important tool that aided in their overall academic success at Purdue (Arnold & Pistilli, 2012).

The use of Course Signals demonstrates the potential for academic analytics. It could certainly be argued that these instructors, armed with data provided by learner analytics, are the most important weapons against student under performance in the classroom (Arnold & Pistilli, 2012) Because of the ability of academic analytics to assess risk early and in real time, the instructors consistently indicate that students are benefitting from knowing how they are really doing in a course, and moreover, understand the importance of completing assignments, and performing well on quizzes and tests (Arnold & Pistilli, 2012). Faculty also say that students tend to be more proactive as a result of the Course Signals interventions (Arnold & Pistilli, 2012).

Ethical Dimension

The positioning of the fourth dimension is by no means meant to assign significance or value to the discussion for data analytics. Rather, concluding with this dimension allows for a richer discussion as the ethical dimension is embedded within each of the previously discussed dimensions. Specifically, this dimension identifies the ethical issues that need to be addressed in the design, development, and implementation of courses, new initiatives, and programs. Issues pertaining to social and political influence; diversity; bias; the digital divide; information accessibility; etiquette; and legal issues, such as privacy, plagiarism, and copyright, are also addressed. The ethical dimension then, addresses issues related to the responsible and ethical use of mined data, as well as the protection and anonymity of human subjects.

Key Issues

The rapid adoption and expansion of learning analytics in the higher education sector has occurred at a faster pace than the consideration of ethical issues surrounding their use (Slade & Prinsloo, 2013; Swenson, 2014) Key ethical issues related to the use of big data and learning analytics are privacy, consent and how data is used, stored, protected and acted upon (Cumbley & Church, 2013; Rubel & Jones, 2016). Of particular concern is the absence of student voices in decision-making concerning learning analytics (Roberts et al., 2016). Students should have an active voice in determining what data is collected about themselves, how it is used and stored, who will have access to the data and how student identities will be protected (Slade & Prinsloo, 2013). While it is important for students to have a voice, it is difficult to determine exactly how that should be handled. As previously noted, there has been an explosion in enrollments in online programs and the amount of data collected is astronomical. It is simply not feasible or realistic to solicit feedback from every student; but there could be a middle ground.

Student Concerns

For starters, students need to be better informed about what learning analytics is and what it means for their data (Roberts et al., 2016). Students need to be educated on what is happening but more importantly *why* it is happening. Therefore, framing the discussions around how the collection and use of the data will be of benefit to the students will go a long way to creating a more informed student body. Once they are better informed, they can be integrated into the decision-making processes. In fact, they may be better able to offer suggestions or identify administrative blind spots that can overall improve the way data is collected, analyzed and used.

At the same time, there is a concern that students may feel these systems actually limit their ability to be move freely through their learning environments (Beattie, Woodley, & Souter, 2014) or that they are always being watched. Netflix and Amazon, two systems that make heavy use of learning analytics, seem like the type of limiting systems students fear in educational contexts. While the recommendations for Netflix and Amazon can be immensely helpful, they can also feel a bit ‘creepy' and seen as insidious to some. For example, a person can be left feeling uncomfortable when they are using Facebook, an ad pops up on the right pane for a shirt that they searched for on Amazon several days prior. Have Facebook or Amazon violated any specific terms? Probably not. However, it does not reduce the uncomfortableness of the situation. The same is potentially true within educational contexts. Students may not be consciously aware that the LMS is capturing so much data on their activities (Sabourin, Kosturko, Fitzgerald, & Mcquiggan, 2015). If an instructor then starts data mining that information to look at relationships and patterns, students may feel just as uncomfortable as individuals targeted by ads within Facebook.

The best way that an instructor can mitigate these feelings is through transparency (Dringus, 2012; Sabourin et al., 2015). The intention of data mining and learning analytics should always be about improving student learning. Thus, the instructor should share with students exactly what they are doing. Explaining that the instructor will be using data collected on student engagement and participation to identify ways to improve the course will likely put many students’ concerns at ease. Also, the students may be willing to provide additional information and context that can better frame the collected data.

Conclusion and Next Steps

Understanding student learning and behavior is a complex process and one that can be significantly aided by data analytics. There is so much collected data that can provide rich context which can in turn inform student learning behaviors and influence learning pathways. However, the implementation and use of learning analytics is not something that can be done quickly or haphazardly, and requires an integrative approach that looks at various data points to create a holistic view and understanding of the learning environment. The raw data is likely already collected in most cases and can be found in LMS or other tools being used in the classroom. The challenge for the researcher is aligning their research questions to the most appropriate mining or analytical approach. Over time, heavily labor-intensive approaches will become significantly more efficient in terms of time. An excellent example of this is in spam filtering – every email system or program has some level of spam filtering and it is significantly improved from where it was just three years ago. Over time, spam filtering will get more and more refined, just as our use of data mining and learning analytics to support instructional uses will improve as well. There are some excellent opportunities for future and new research in the use of learning analytics in education and text mining will feature prominently in these developments.

Few can argue that data analytics does not yield useful and exciting results. However, as the amount of data that is being studied increases, so too does the amount of work it can take to fully analyze that information. Furthermore, time is a prized limited resource, and is therefore among the most significant challenges to text mining. In several of the studies highlighted in this chapter (e.g. Ai et al, 2010; Ezen-Can, 2015) there was a tremendous amount of data that needed to be collected and analyzed. For many this can be a limiting factor.

The decisions required for dealing with the rapid changes within higher education are complex and many are made without recourse to vast data sources that have been generated but are not available to those entrusted to make relevant and timely choices (Daniel, 2015). There is still a divide between those who know how to extract data and what data are available; and those who know what data are required and how it would best be used, all of which make collaboration difficult (Daniel, 2015). There are no easy options in developing policies and systems that address the intersecting and conflicting attitudes held by students however, the starting point needs to be engaging students in the decision making process (Roberts et al., 2016).

Questions for Class Discussions

1. How should universities inform students about the way their data is being used?
2. Could engaging students in the decision-making process possibly have a negative impact?
3. How can universities better analyze and utilize the student data that is being collected?

4. Are there opportunities for universities to collect data on instructors that would be useful for students? If so, what would that process look like? What type of data would be collected?
5. In what ways could learning analytics be used to provide feedback that is even more personalized for students?

References

Ai, H., Sionti, M., Wang, Y.-C., & Rose, C. P. (2010). Finding transactive contributions in whole group classroom discussions. In *Proceedings of the 2010 International Conference of the Learning Sciences* (Vol. 1, pp. 976–983).

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12* (pp. 267–270). New York, New York, USA: ACM Press. http://doi.org/10.1145/2330601.2330666

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, *113*, 177–194. http://doi.org/10.1016/j.compedu.2017.05.007

Baker, R. (2010). Data Mining. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd ed., Vol. 33, pp. 112–118). Elsevier. http://doi.org/10.1016/B978-0-08-044894-7.01318-X

Baker, R. (2013). Learning, schooling, and data analytics. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on Innovations in Learning* (pp. 179–190). Philadelphia, PA: Center on Innovations in Learning, Temple University.

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3–16.

Beattie, S., Woodley, C., & Souter, K. (2014). Creepy analytics and learner data rights. In B. Hegarty, J. McDonald, & S.-K. Loke (Eds.), *Rhetoric and Reality: Critical Perspectives on Educational Technology (ascilite 2014)* (pp. 421–425). Dunedin, NZ.

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational Data Mining and Learning Analytics: Applications to Constructionist Research. *Technology, Knowledge and Learning*, *19*, 205–220. http://doi.org/10.1007/s10758-014-9223-7

Brown, A., Cowan, J., & Green, T. (2016). Faculty productivity: Using social media and measuring its impact. Retrieved October 7, 2017, from https://er.educause.edu/articles/2016/5/faculty-productivity-using-social-media-and-measuring-its-impact

Cumbley, R., & Church, P. (2013). Is "Big Data" creepy? *Computer Law & Security Review*, *29*(5), 601–609. http://doi.org/10.1016/j.clsr.2013.07.007

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, *46*(5), 904–920. http://doi.org/10.1111/bjet.12230

de Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., … Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175–1188. http://doi.org/10.1111/bjet.12212

Dede, C., Ho, A., & Mitros, P. (2016). Big data analysis in higher education: Promises and pitfalls. *Educause Review*, *51*(5), 23–34.

Dringus, L. P. (2012). Learning analytics considered harmful. *Journal of Asynchronous Learning Networks*, *16*(3), 87–101.

Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15* (pp. 146–150). New York, New York, USA: ACM Press. http://doi.org/10.1145/2723576.2723589

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. http://doi.org/10.1007/s11528-014-0822-x

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, *15*(3), 42–57.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher*, *43*(6), 293–303. http://doi.org/10.3102/0013189X14544542

Gupta, V., & Lehal, G. S. (2009). A survey of texting mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, *1*(1), 60–76. http://doi.org/10.4304/jetwi.1.1.60-76

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and leavers: Early-Career teacher effectiveness and attrition. *Educational Researcher*, *40*(6), 271–280. http://doi.org/10.3102/0013189X11419042

Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, *13*(1), 61–72. http://doi.org/10.2478/cait-2013-0006

Khan, B. (2001). A framework for web-based learning. In B. Khan (Ed.), *Web-based training* (pp. 75–98). Englewood Cliffs, NJ: Educational Technology Publications, Inc.

Khan, B. (2005). *Managing E-Learning: Design, delivery, implementation and evaluation*. Hershey, PA: Information Science Publishing.

Kruse, A., & Pongsajapan, R. (2012). Student-centered learning analytics. *CNDLS Thought Papers*, 1–9. http://doi.org/10.1007/978-1-4419-1428-6_173

Leong, C. K., Lee, Y. H., & Mak, W. K. (2012). Mining sentiments in SMS texts for teaching evaluation. *Expert Systems with Applications*, *39*(3), 2584–2589. http://doi.org/10.1016/j.eswa.2011.08.113

Martin, F., & Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice*, *13*(3), 1–20. http://doi.org/10.1177/0047239516656369

Papamitsiou, Z., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society*, *17*(4), 49–64.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450. http://doi.org/10.1111/bjet.12152

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20. http://doi.org/10.24059/olj.v16i3.267

Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student attitudes toward learning analytics in higher education: “The Fitbit version of the learning world.” *Frontiers in Psychology*, *7*, 1–11. http://doi.org/10.3389/fpsyg.2016.01959

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, *32*(2), 143–159. http://doi.org/10.1080/01972243.2016.1130502

Sabourin, J., Kosturko, L., Fitzgerald, C., & Mcquiggan, S. (2015). Student Privacy and Educational Data Mining : Perspectives from Industry. *Proceedings of the 8th International Conference on Educational Data Mining*, 164–170.

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. http://doi.org/10.1177/0002764213498851

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. http://doi.org/10.1177/0002764213479366

Swenson, J. (2014). Establishing an ethical literacy for learning analytics. In *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK ’14* (pp. 246–250). Indianapolis, Indiana: ACM Press. http://doi.org/10.1145/2567574.2567613

Wong, J.-S., Pursel, B., Divinsky, A., & Jansen, B. J. (2015). An analysis of MOOC discussion forum interactions from the most active users. In N. Agarwal, K. Xu, & N. Osgood (Eds.), *Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 452–457). Switzerland: Springer International Publishing. http://doi.org/10.1007/978-3-319-16268-3_58

Yang, L. (2013). Big Data analytics: What’s the big deal? Retrieved September 22, 2017, from http://knowledge.ckgsb.edu.cn/2013/12/30/technology/big-data-analytics-whats-big-deal/