Learning Analytics For Engineering Education

DR. LESLEY S. J. FARMER
Advanced Studies in Education and Counseling Department
California State University Long Beach
1250 Bellflower Blvd., Long Beach CA 90840
UNITED STATES OF AMERICA
Lesley.Farmer@csulb.edu   http://www.csulb.edu/~lfarmer

Abstract: - Engineering faculty should take advantage of data analytics to improve their curriculum, program, and student success. As such, faculty need to strategically conduct the entire data process: knowing the right questions to ask, determining the relevant data to collection, choosing the appropriate instruments to collect those data, analyzing that data, recommending appropriate actions, implementing them, and evaluating the implementation. Furthermore, students need to learn how to analyze data as part of their professional repertoire of design and development intellectual toolkits. To that end, learning analytics practices are detailed in this paper. Engineering education constitute the contextual focus.

Key-Words: - learning analytics, data analytics, data, curriculum, instructional design

1 Introduction

In today's data-driven society, numbers and other evidence abound, including in higher education. However, data by itself is not very useful or even informative. Engineering faculty need to strategically conduct the entire data process: knowing the right questions to ask, determining the relevant data to collection, choosing the appropriate instruments to collect those data, analyzing that data, recommending appropriate actions, implementing them, and evaluating the implementation. Furthermore, their students need to learn how to analyze data as part of their professional repertoire of design and development intelletual toolkit.

One of the main sticking points in such data procedures appears to be the analysis itself. How do manager make sense of the data? What do the story behind the data tell? While accurately representing the data is a good first step, explaining the data and making inferences about the findings is necessary in order to make logical, feasible recommendations for action.

Within this set of factors, engineering educators should also apply data analytics to examining their own curricular practice and student learning. To that end, learning analytics practices are detailed in this paper. Engineering education constitute the contextual focus.

2 Data Analytics

Data analytics may be defined as a scientific and systematic approach to examine raw data in order to draw valid conclusions about them. Data analytics includes extracting and organizing raw data, determining the appropriate statistics, applying those statistics to the resultant data, analyzing and interpreting the results, and making recommendations based on the data analysis. Data analysis is one step within data analytics, and focuses on evaluating the data logically; it is more procedural than conceptual.

2.1 Data Analytics in Engineering Education

Engineers constantly deal with data as they iteratively solve engineering programs and design effective and efficient products. At each step in their design process, engineers assess the feasibility, accuracy, thoroughness, and effectiveness of their efforts. The process typically depends on technical skill and insight. The 2016 national ABET program standards include student outcome criteria that address data analytics: applying mathematical knowledge, designing and conducting experiments, and analying and interpreting data [1]. Additionally, Lindsay Montanari, associate director of the Analytics MS program at Northwestern University's school of engineering, asserted "companies were looking for graduates with not only the technical skills, but also the business skills" [2].

In their 2011 study of data mining courses in graduate computer science and engineering programs, Safer, Farmer and Chuk found that nearly half of those program offered a data mining course. It should be noted that of the data mining software mentioned, 18% listed Matlab, and half listed a programming application. SAS mentioned for only 3% of the time, and Python was listed only 2% of the time. Likewise, the courses tended to use textbooks on machine learning and neural networks more than on data mining itself. Computer science and engineering faculty published an average of 9.4 articles per year per university, rising nearly fifty percent over five years [3].

2.2 Learning Analytics

Learning analytics is a branch of data analytics that focuses on student learning. Long and Siemans define learning analytics as Long and Siemens define learning analytics as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" [4]. It examines data about students and their learning enviroment, the latter of which may include the curriculum, instruction, resources, the physical and virtual learning environment, and other possible contextual factors. Analytics not only look at the individual factors but also their interaction and relationship; within a course, these factors can be clustered into the following categories: student-information (e.g., learning management system and its resources), student-student (sometimes referred to as social learning analytics), and student-teacher.

Within that umbrella educational data mining employs computerized methods to identify possible trends within large educational data sets [5]. As such, then, learning analytics may be applied at several levels: individual student or instructor, single or multiple learning activities, resources, course, program, student or faculty group, university, system, and geographic region. Sometimes the term "academic analytics" is used to focus on the business intelligence side of education.

Learning analytics facilitates improved educational experiences: increased learning, more effective learning, deeper knowledge and skills, greater equity in learning, higher-quality curriculum and instruction, and improved programs and institutions. Learning analytics can help educators model student behavior and learning systems, predict factors for student success, optimize interventions to impact learning positively, and generally plan educational experiences and conditions for learning more systematically.

It should be noted that learning analytics, as in other educational data analytis, cannot be reduced to a simple algorithm or statistic. Education is an ill-structured science, with many variables that cannot be controlled. Educational institutions rarely can control a student's life outside the institution, be it living conditions, human relations, internal dispositions, or life situations. Nevertheless, particularly for large populations over time, learning analytics can reveal important factors that can lead to improved conditions for learning.

3 Data Factors

3.1 Purposeful Data

"Measure what you value" serves as a guiding mantra for learning analytics. Data analytics begins with identifying its purpose: to improve the organization and its management. In order to do this, managers first need to assess the current status of the organization and its stakeholders, relative to the organization's mission and goals. These baseline data can then serve as the basis for identifying areas for improvement and growth. For learning analytics, management consists of educators' management of the learning sytsem.

Each educational system generates data, be it consciously gathered, such as in a student survey, or automatically such as enrollment counts or website hits. Building on the above systems approach, the following data might be generated or gathered, within the concrete context of engineering education.

Input:

Learning management system: online hits (frequency, date/time, length of time online), discussion (quantity, quality), quizes (answers)

Resources: volume, currency, subject matter (quantity, quality), formats (quantity, quality), cost of items, accessibility

Space: physical dimensions, accessibility, appearance and organization), furniture (quantity, quality, accessibility), utilities (connectivity, quality, cost), operating hours (timing, quantity)

Personnel: staff (number, demographics, hours, salary, position, retention), volunteers (number, demographics, hours, productivity, retention), consultants (number, demographics, hours, salary)

Finances: funding sources and amounts, allocation

Activities:

Curriculum development: selection, organization

Instruction (number and demographics of instructors and target audience, format, location, collaboration quantity and quality, mode of delivery, content, scheduling, length per instructional period); creation of learning aids such as bibliographies (quantity, quality, process time); personnel training (number and demographics of instructors and target audience, format, location, collaboration quantity and quality, mode of delivery, content, scheduling, length per instructional period)

Facilities: use of space, scheduling

Fiscal: budgeting, accounting, ordering (productivity, process time, efficiency, accuracy, failure rate)

Output:

Student participation: quantity, quality, degree of satisfaction, repeated participation over time

Student products: quantity, quality

Facilities use quantity, quality, timing, wear and improvement

Outcomes:

Student achievement: assignment assessment, grades, awards and other public recognition, employment, promotion, increased income

Program: retention rate, graduation rate

In sum, thousands of data points may exist; engineering educators need to identify which data are needed in order to ascertain the gap between the current and target situation. In those cases where data are not collected, managers need to determine the data collection method. Some typical means follow, noting their strengths and weaknesses:

- Observation: provides open-ended, direct evidence in a natural setting. Individuals bring their own biases and may see different things so training and consistent results are needed; observation captures data only for that instant.
- Individual interview: provides open-ended, interactive, in-depth data. Labor-intensive and time-consuming, the data are only as accurate as the questions being asked and individual being interviewed so this approach requires training; language barriers may exist.
- Focus group: provides open-ended, interactive, in-depth data and group dynamics. Data may be skewed or missing because of group norming; this approach requires training; language barriers may exist as well.
- Online interview: provides open-ended, interactive, in-depth data that is space-dependent and less intimidating. Interviewers need training, and data may be biased because of access (non-ownership or unlisted) and language barriers.
- Content analysis: provides unobtrusive data that can be repurposed. Confidentiality may limit access or application; data may need to be contextualized.

3.2 Systematic Data

One of the main barriers to effective data analytics is the lack of a systematic or strategic approach. At the same time that engineering educators may be flooded with data, they may find data to be meaningless, decontextualized, or disconnected from the rest of the system – and comparable programs.

Internally, engineering educators need to identify the different functions and systems within the organization, and how those functions interact. An organizational flowchart provides a visual way to understand the relationships. Then managers can identify what data are associated with each function and each relationship. What data already exist, and what new data need to be collected? For instance, a core function of engineering education is design process. As readings and other resources are added to the learning management system, several kinds of data are generated. In terms of course delivery, learning management systems typically generate data automatically such as course access behaviors, course partication input, quiz and survey data, and grades.

Engineering educators need to look outside of the organization's internal system to be strategic in analyzing data, especially in terms of academic analytics. For instance, if a thousand users access the program's website, what conclusions can the engineering educator make? For a very specialized function, that number may reflect good market penetration, but engineering educator will not know unless they can benchmark that number with competitor websites. How long did the user stay on a page? One conclusion might be that the user quickly found the information needed, but it might be concluded that the user was not interested in the page's content, or disliked its interface or appearance. Engineering educators need to ascertain the reason for student behavior, so in this case, web accessibility studies would need to be conducted to address those conjectures. Both of these issues demonstrate the need for a strategic approach, one that involves interacting with the external environment of the organization, be it students, comparable programs, or the university at large.

4 Data Analysis

Data are extracted and structured, and qualitative and quantitative techniques are used to identify and analyze patterns. Qualitative data analysis usually works with non-numerica data (e..g., interviews, objects, observations), and codes those data to interpret and explain the data. Quantitive data analysis work with numerical data.

As noted above, a major obstacle in using data is the analysis itself. Several factors may account for this phenomenon: engineering educator's lack of statistical background , statistician's lack of organizational and management knowledge, lack of technical or statistical personnel, differing status between educator and technical (e.g., statistician) personnel, perceived disconnect between data and adminstrative practices, perceived lack of available time, perceived poor ROI (return on investment) of data analytics, poor data quality, lack of integration of data with existing infrastructure, insufficient systems to process data efficiently, perceived lack of data analytics tools (or lack of knowledge about such products), other higher-priority administrative demands [6][7].

4.1 Preparing the Data

For engineering educators to analyze data meaningfully, they have to make sure the data are well structured and "clean": that is, free of typos, missing values, and extreme values. Typically, quantitative data are organized into a spreadsheet or statistical program (e.g., MiniTab, SPSS), with the variables (e.g., grades, discussion, hits) listed one per column, and the responses or observations (e.g., each reference transaction, training session, staff member, item processed) listed one per row. These variables should be defined at the beginning of the data analtyics process as engineering educators define the issue to examine, and the possible factors or variables that impact the issue.

Each variable needs to be clearly defined in terms of its meaning (e.g., graduation status), its format (e.g., how data is written), and type of variable (e.g., numeric, string of characters). To facilitate data analysis, responses might be coded, such as level of education (1=grade school, 2=high school, 3=some college, 4=bachelor's degree, 5=post-graduate). Sometimes a new varible might be generated from existing data, such as time per process (time online multiplied by number of hours to do process).

4.2 Data Representation

A starting point in data analytics is descriptive statistics; common statistics include minimum and maximum, mean (average), median (middle value), quartiles, and standard deviations (which measure variability). Visualizing the data often helps engineering educators make sense of the findings. Usual graphs include histograms or bar graphs to show frequencies, box plots to show quartiles and range, and pie charts to show percentages of a whole. These graphs can give managers an idea of the scope of the findings, and help them to explore if two sets of findings reflect similar or different populations. Visualizing the data also helps communicate findings to stakeholders.

As much as possible, data should be disaggregated by demographics such as sex, age, ethnicity, and program status in order to help identify at-risk groups; if possible, data can also be disaggregated by time frame or subject matter. For example, females tend to discuss less online so interventions should be custom-designed to motivate and help that group. Another way to disaggregate data is by quartiles or other score rankings; one cost-effective practice is to focus on those groups who almost meet a standard because a specifically targeted intervention may be relatively easy to implement and result in a significant return on investment.

4.3 Matching the Statistical Test to the Type of Data

To optimize the use of data, engineering educators should apply inferential statistics in order to make or infer generalizations about significantly large populations based on sampling. For example, analysis tries to find correlations between two variables, such as number of hours on a course website opening hours and project grades.

If the number of responses or observations is limited such that engineering educators cannot assume that the population is not normally distributed, which is more likely to occur when the number of responses is limited, then non-parametric statistics need to be applied. Large data sets tend to use parametric statistics.

The most important statistical consideration is the characteristic of the derived numbers; misaligning a statistical method with number

property causes misleading conclusions. The chief error is ascribing mathematical equations to emotions (e.g., one person is 2.5 times as satisfied as another person). Data are usually categorized as nominal, ordinal, internal, and ratio scalar.
Data may also be distinguished as discrete (whole numbers such as the number of user) or continuous (analogue such as length). Most numbers in assessment are discrete ordinal or interval kinds.

The following statistical tests represent the most common data analytics techniques.

- Correlation analysis. This statistical method quantifies linear relationships between two variables. The analysis generates a correlation coefficient r that tells the strength of the relationship and whether that relationship is positive or negative. For example, what is the correlation between the number of discussion entries and grades? This analysis of continuous values uses a Pearson correlation measure; a chi square measure is used for categorical data (e.g., gender).
- Regression analysis. This statistical method tries to generate a model (i.e., a line or curve that fits the data best) that shows the relationship between pairs of numerical data. Unlike correlation analysis, the regression line may be nonlinear. Several variables might affect one dependent variable, which would use multiple regression. For instance, a regression analysis might inform engineering educators about optimum length of times for instruction.
- ANOVA (analysis of variance). Variability of data determines if greater differences in data exist between groups than within a group. For instance, do graduate students get more relevant website "hits" than freshmen; does more variance occur within or across subject domains or experience?
- Cluster analysis. This statistical method uses distances between variables to group observations together. Those with smaller distances between them are assumed to be similar, so looking closer at the individual clusters can potentially determine important characteristics. For example, how might information behaviors cluster for successful retrieval of information?
- Factor analysis. This method identifies main efforts (that his, response by a change in tehf ctor's level) when several factors, or variables eists. It can test all possible combinations of the factors and their respective levels (e.g., high, medium, low use) to determine possible interactions.

4.4 Student Learning Analytics

Most engineering curriculum includes data analytics as part of design processes and experimentation. Several textbooks explain statistics such as data modeling, failure analysis, and quality control. However, data analytics can apply to students' own learning. While educators are likely to have greater access to data than students, data analytics should incorporate student perspectives. For instance, engineering educators can intrerview students to understand the reason for certain findings (e.g., patterns in group work related to grades). Furthermore, students should analyze data about their own performance in order to optimize their learning (e.g., time of day that they access the learning management system). Students can also review their peers' learning behaviors in order to identify trends in learning and to suggest interventions for improved learning.

5 Conclusion

Learning analytics inform engineering educators as they strive to provide optimum learning experiences for all their students. Especially as faculty can identify significant factors for student success, they can develop targeted interventions to optimize student learning. Faculty can also identify those institutional factors that impact instruction and learning, and use that evidence to influence administration to address those issues. Furthermore, engineering educators can involve their students in such processes, and facilitate students' own efforts in analyzing their learning experiences, which can inform their own efforts and bring them control of their learning.

Reference:

[1] ABET, *Criteria for Accrediting Engineering Programs, 2016-2017*, ABET, 2016.
[2] Tableau, *The State of Data Education in 2016*, Tableau, 2016, p. 6.
[3] Safer, A., Farmer, L., & Chuk, E., Data Mining Technology across Academic Disciplines, *Intelligent Information Management*, Vol.*3*, No.2, 2011, pp. 43-48.
[4] Long, P., & Siemens, G., Penetrating the Fog: Analytics in Learning and Education, *Educause Review*, Vol.46, No.5, 2011, pp. 31-40.

[5] Romero.,C., & Ventura, S., Data Mining in Education, *WIREs Data Mining and Knowledge Discovery*, Vol.3, 2013, pp. 12-27.
[6] Finos, R., Wikibon Big Data Analytics Survey, SiliconANGLE Media, 2015.
[7] Henschen, D., *2015 analytics and IB Survey*, UBM Tech, 2014.