The Seventh International Conference on eLearning (eLearning-2016), 29 – 30 September 2016, Belgrade, Serbia

LEARNING ANALYTICS: WHAT DO COMPANIES KNOW ABOUT OUR KIDS (AND WE DON'T)?

KRISTIJAN ZIMMER
University of Zagreb, Faculty of Organization and Informatics, Croatia, Kristijan.Zimmer@foi.hr

Abstract: *The article provides an introduction to the increasingly important field of Learning Analytics (LA), new opportunities as well as threats from not that unlikely places. As we gather students' learning-related data, we should be more able to help them by providing them with means of understanding their progress as well as offering them "early warning system" for potential failure. If such systems offer other useful features such as problem reporting tools or achievement history, they bring additional new value to education in terms of better communication and motivation.*

However, beside students and teachers, many other people usually have access to some aspects of that data and their analytical representation. Some people should have an access but lack knowledge of interpretation and/or are not aware of crucial ethical and privacy issues. On the other hand, some who do, clearly shouldn't.

The article focuses on companies which provide IT systems for different purposes related to education. Should we entrust them with our kids' grades, homeworks, records of daily activities, communication with teachers and peers, especially if such services are cloud-based? Can this be used for "silent" future job profiling proposes? Can our kids be blackmailed some day? The article covers potential future(s), from benevolent ones to some extremely Orwellian.

Keywords: *Learning analytics, educational information systems, privacy, ethics*

1. INTRODUCTION

Learning Analytics has become increasingly important field of educational research in the last decade. There is no single definition for the term, however one of the most frequently used is the one by professor Erik Duval from the KU Leuven in Belgium defining learning analytics as "collecting traces that learners leave behind and using those traces to improve learning" [1].

Other definitions include Horizon 2016 Report [2] which describes learning analytics as "an educational application of web analytics aimed at learner profiling, a process of gathering and analysing details of individual student interactions in online learning activities."

In many scientific publications learning analytics is perceived as a key technology for the improvement of education. It is also referred to as "big data" applied to education. Beginnings of this scientific field come partly from the commercial sector that has used analyses of consumer activities to identify key consumer trends.

Educational data mining (EDM) uses similar techniques of consumer data mining, only in the context of education on all levels.

Typical phases of educational data mining projects consist of preparation of data through the process known as ETL (extract, transform, load), followed by mathematical analyses, visualisations (often including interactive dashboards) and actions triggered by the results.
Advanced application of learning analytics include mathematical modelling that can be grouped in 4 types of analyses: descriptive, diagnostic, predictive and prescriptive.

Descriptive analysis aims to explain what is happening, diagnostic helps in understanding why certain things happened, predictive helps in prediction what will happen in the future, while prescriptive suggests the best course of action to take to optimise business outcomes. (3)

In job profiling there are elements of all four types of analysis, however with focus on the first two, especially with the goal of classification of candidates.

2. CORPORATE VIEWPOINT

What would be the incentive for the companies providing information systems with educational data to analyse them and by doing so as well analysing their clients – the students? Would they do it for purposes other than sole interests of their clients?

In typical corporations, set of values and resulting business ethics are a bit different than in education. "If a company's purpose is to maximise shareholder returns, then sacrificing profits to other concerns is a violation of its fiduciary responsibility" [4]. This in many aspects extreme viewpoint is confronted with more moderate and modern "corporate social responsibility" set of values, emphasizing

more holistic view which includes protecting interests of customers, other citizens, local communities, employees of companies in their supply chain, nature preservation, etc.

Analysis of the existing and perspective clients is a key business activity for any company, aimed at providing business goals such as:

- understanding clients' needs;
- classification of clients based on historical data;
- predicting clients' future decisions;
- maximizing sales and client retention opportunities;
- minimizing lost sales opportunities.

In cases when the clients are educational institutions and their students, it is generally less known to the public how the companies providing educational information system make use of potentially very interesting future job profiling value of this data. Companies that possess such data could potentially use them for their own recruitment needs or could create a business model to offer services to another legal entities.

3. EDUCATIONAL INFORMATION SYSTEMS, DATA AND POTENTIAL ANALYTICAL OUTCOMES

Today, there are numerous information systems used in all levels of education, from primary to higher. In this article only information systems containing educational data about students are presented. Those information can be used to analyse student's success, performance, behavioural patterns, interests, social connections, etc.

The following is the list of types of commonly used educational information system, types of data they contain potentially relevant to job profiling and potential outcomes of analyses.

Information system: **Student information system (SIS)**
Relevant data: courses, teaching staff, students, enrolments, mid-term and final grades, timetable, attendance, financial data (tuition, loans), socio-economic data (address, previous education).
Potential analytical outcomes: success, progress, recoveries from failures, learning outcomes / competencies, motivation, interests, health, financial status, socio-economic status.

Information system: **Learning management system (LMS)**
Relevant data: courses, teaching staff, students, enrolments, learning content (read/viewed, created, edited), lesson progress, tasks/assignments, communication (initiated, replied), testing and self-testing results, polls.
Potential analytical outcomes: success, progress, communication skills, social skills, promptness, punctuality, interests, opinions on selected topics.

Information system: **Productivity (office)**
Relevant data: learning content (read/viewed, created, edited), social interactions (comments), tasks / assignments
Potential analytical outcomes: success, progress, punctuality, interests, opinions on selected topics.

Information system: **Team / project management system**
Relevant data: tasks / assignments, milestones, social connections, communication (initiated, replied).
Potential analytical outcomes: success, progress, communication skills, promptness, project management skills, leadership skills, punctuality, interests, opinions on selected topics.

Information system: **Learning object repository (LOR)**
Relevant data: learning content (read/viewed, created, edited), social connections, communication (initiated, replied), social interactions (likes, comments recommendations, shares).
Potential analytical outcomes: interests, opinions on selected topics, communication skills.

Information system: **Social networks**
Relevant data: social connections, communication (initiated, replied), social interactions (likes, comments recommendations, shares).
Potential analytical outcomes: social skills, communication skills, interests, opinions on selected topics.

Information system: **E-mail system**
Relevant data: social connections, communication (initiated, replied).
Potential analytical outcomes: social skills, communication skills, promptness, interests, opinions on selected topics.

Information system: **Video on Demand (VoD)**
Relevant data: learning content (read/viewed, created, edited), social interactions (likes, comments recommendations, shares).
Potential analytical outcomes: interests, opinions on selected topics.

Information system: **Learning record store (LRS)**
Relevant data: learning content (read/viewed, created, edited), communication (initiated, replied), social interactions (likes, comments recommendations, shares), communication (initiated, replied), testing and self-testing results, polls.
Potential analytical outcomes: success, progress, interests, communication skills, promptness, opinions on selected topics.

Information system: **Classroom management system with Mobile Device Management (MDM)**
Relevant data: courses, teaching staff, students, enrolments, learning content (read/viewed), screens (teaching staff and students), communication, whiteboards, testing and self-testing results, polls, time management log, device control log (lock / unlock, application launch), device's physical location log.

Potential analytical outcomes: success, progress, interests, communication skills, promptness, opinions on selected topics, social interactions in physical space.

Information system: **Educational mobile / web applications**
Relevant data: learning content (read/viewed, created, edited), communication (initiated, replied), testing and self-testing results, polls.
Potential analytical outcomes: success, progress, interests, communication skills, opinions on selected topics.

Information system: **Interactive whiteboards**
Relevant data: whiteboard content.
Potential analytical outcomes: interests.

Information system: **Access control / management**
Relevant data: students' access log for lecture rooms and other physical and virtual learning spaces.
Potential analytical outcomes: interests, motivation, punctuality, health.

Information system: **Assessment software**
Relevant data: testing and self-testing results.
Potential analytical outcomes: success, progress, interests.

As this overview of most commonly used educational information systems shows, many systems contain sensitive, personal data: grades and other elements of success, abilities to recover from failure and ability rarely/never to enter critical situations needing recovery, amount, quality and promptness of communication and contributions, motivation at certain points of schooling, interests and potential health issues visible in longer periods of absence.

4. AUTHENTICATION TYPES AND PRIVACY ISSUES

All potential privacy issues arising from availability of such data to commercial companies "multiply" with how closely this data can be tied to a certain person and how easily is to gather and integrate such data.

National and international authentication ID schemes, with initial authentication of the user from a trusted, official source are the worst in that regard. On the other hand, such schemes are the best for introducing intelligent government and other services to its citizens.

Many global (cloud) service providers offer their authentication schemes. Examples of such companies are Google, Microsoft, Apple and Yahoo. They usually rely on user's e-mail address used in login creation process. This e-mail is generally unreliable and may not reveal genuine name or other credible elements of identity. Governmental ID services offer reliable identification of the user and usually tie the username with national or international citizen ID number. If this information is passed on to the commercial provider at any stage, privacy is threatened and needs to be managed and monitored with maximum care.

There are several levels of confidence when pairing a certain person to a user of an educational online service.

- **Anonymous access**
  In this case there are little privacy issues, since users are anonymous. Such services are nowadays very rare. One possibility of such service to remember the returning user is by means of browser cookies. Using another browser would make service forgetting the entire history for the returning user, making the service unintelligent and in many cases practically useless.

- **Username not tied to the real, verified name**
  This authentication in which user chooses his/hers own username and password, without reference to an existing e-mail account or other trusted authentication scheme is a candidate for the best option concerning privacy. It is rare, since service provider companies wish to have a reference to a more concrete user identity in case of issues concerning illegal content or activity, rather than just a timestamp and the IP address, that can be hidden behind multiple VPN's or anonymity networks such as Tor [5].

- **Real, verified name, connected to school-level ID scheme**
  In this authentication scheme students are given username/passwords pairs by the schools authority and part of this information is sent to the commercial provider as part of the SSO (single sign-on) procedure, usually with unchanged user ID. Full name and e-mail address are usually sent along with user ID to enable correct addressing and e-mail communication with users. Such approach is far from ideal from the privacy point of view, but in a global scenario would require collecting user data from many individual schools by the global providers to create a relevant database for a global commercial use. With educational services markets already segmented, making a global database is practically impossible without radical corporate acquisitions and mergers.

- **Real, verified name, integrated with a permanent regional / national / international ID scheme**
  This authentication scheme presents potentially the most dangerous combination, concerning the privacy. For a global provider, this would enable creating global database of users and a basis for all kinds of analytics, potentially usable for job profiling purposes.

5. ETHICAL STANDARDS AND CORPORATE PRACTICES

All this has a lesser negative impact on privacy if ethical standards are applied by all parties, including commercial companies.

One of the relatively new ethical recommendation comes in the form of 8-step guideline known as "DELICATE" [6]:

1. *Determination: Decide on the purpose of learning analytics for your institution;*

2. *Explain: Define the scope of data collection and usage;*
3. *Legitimate: Explain how you operate within the legal frameworks, refer to the essential legislation;*
4. *Involve: Talk to stakeholders and give assurances about the data distribution and use;*
5. *Consent: Seek consent through clear consent questions;*
6. *Anonymise: De-identify individuals as much as possible;*
7. *Technical aspects: Monitor who has access to data, especially in areas with high staff turn-over;*
8. *External partners: Make sure externals provide highest data security standards.*

Another important issue are legal obligations and commitments arising from the contracts, "terms of use" documents and end-user licenses. Nowadays, most commercial providers dedicate themselves in protecting user data from the "third parties", while explicitly stating intention of using data themselves for the purpose of making their software and services better and more useful for their users. How exactly this goal will be achieved is often not explicitly stated. This almost always include possibility of automated data gathering and analytics, since knowing clients is a key to success and is integrated in the business practice.

Microsoft claims to use clients' data "*only to provide customer the online services including purposes compatible with providing those services. For example, (Microsoft) may use customer data to provide a personalised experience, improve service reliability, combat spam or other malware, or improve features and functionality of the Online Services. Microsoft will not use customer data or derive information from it for any advertising or similar commercial purposes.*" [7]

Some companies such as Samsung inform and ask the users of their "Samsung Smart School" platform to accept the extensive monitoring of the use of their services [8]:
"*In addition to the data you provide, we may collect information about your use of our services via software on your device and other means. For example, we may collect:*

- *Information about the product - hardware model, IMEI number and other unique device identifiers, MAC address, IP address, operating system version and device settings that you use to access the services.*
- *Information on the application - such as the time and duration of your use of our service, the search terms you enter through our services and any information stored in cookies we set up on your device.*
- *The location data - such as the GPS signal of your device or data about WiFi access points nearby and repeaters that we would be able to transfer when using certain our service.*
- *Voice details - such as recordings of your voice that we record (and possibly store it on our servers) when you use voice commands to manage our service. (Please note that we work with the third party provider who provides speech-to-text on our behalf. This service can receive and store certain voice commands.)*
- *Other information about your use of our services, such as applications you use, websites you visit and how you interact with content that is offered by us."*

The nature and type of data collected data strongly suggests systematic use of analytics, including third-party companies. In the chapter on how Samsung uses collected data the focus is on making products and services better as well as marketing purposes, without mentioning user job profiling. As proposed in "Conclusions and recommendations", standardisations of contract articles regarding privacy and data analyses, EULA's and terms of use should guarantee a global acceptance of privacy policies and standards.

The goal of global market domination for an educational provider means that it should aim to sign as many contracts with schools or counties/states as possible. As job market becomes increasingly global, the pressure to use data for job profiling rises. To increase the market share, companies try to acquire or merge with other companies.

In such processes several scenarios are possible, including:

- Global ID and services provider acquiring educational information system provider(s) (and vice versa less likely);
- Global career/job site acquiring educational information systems provider(s) (and vice versa less likely);
- Global ID and services provider acquiring a global career/job site (and vice versa less likely).

In June 2016 Microsoft, as a global ID and services as well as educational information systems provider [9] acquired world's leading career/job site LinkedIn for 26.2 billion USD [10]. This potentially enables Microsoft to provide its new clients – job recruiting companies using LinkedIn - with services based upon data collected from current school / university students, giving the company an important advantage for its own recruitment process as well.

In the past there have been unsuccessful learning analytics projects, with InBloom case [11] perhaps being the most famous for its failure in communicating project goals with key stakeholders such as parents, failure in proving collected data is secure and lack of opt out possibility.

Good examples of innovative business models involving some aspects of learning analytics could be found in the case of Stanford's massive open online course "CS221: Introduction to Artificial Intelligence" [12] held in 2011. After 160.000 interested students enrolled and 20.000 successfully completed the course, some of the leading tech companies got interested to hire the most successful students. Course authors from Stanford University Sebastian Thrun and Peter Norvig quickly built the business model charging the companies for that information, but first sending an e-mail to the best 1.000 candidates asking for permission to pass on their contact

information to the companies such as Google, interested in their employment. Majority of candidates answered positively, as such business model was both ethical and beneficial for all interested sides.

Later, Udacity [13] (co-founded by Thrun) built on that experience a new, revolutionary model of online education called "Nanodegree Plus" [14], promising to return tuition to students of their niche specialisation online courses if they don't get a job in 6 months following the successfully completed course.

6. CONCLUSIONS AND RECOMMENDATIONS

"Digitalization" of education and related administrative processes has led to appearance of vast quantities of educational data. Learning analytics can help in making such data useful primarily for students, but also for teachers, school and university administration bodies as well as governmental bodies responsible for education on different levels.

Companies that offer educational information systems and educational services collect all sorts of educational data based on their users' educational activities. Such data can be used to create useful educational tools such as personal analytical early warning systems for potential difficulties and upcoming failures, but can also be used for job profiling purposes in the future, making learning environment extremely hostile and unsafe. Market logic is that companies which would employ such practice would be rewarded and gain competitive advantage. Therefore, regulation, monitoring and acceptance of ethical standards are absolutely necessary.

To holistically approach such challenges several initiatives should be undertaken, some of which may include:

1. Widespread application of ethical standards and the best practices of protecting privacy and related data lifecycle (including destruction) in the contracts, terms of use documents and EULAs (end user license agreements) by all involved entities, especially commercial companies that provide educational information systems and related services as well as their employees.

2. Widespread adoption of opt out possibility for students not wanting to be represented in information systems with their real personal information, in situations where opt out of an information system is generally not possible.

3. Widespread adoption of learning analytics in helping students, teaching staff and schools' / universities' management to help them in education and administration. Data driven decision making processes should be implemented at all levels where possible, in an ethical way.

Failure to make use of learning analytics is a failure of management of educational institutions and a good predictor for an educational institution to become obsolete in the globalised educational market for a simple reason for not helping its clients, the students, the best it can.

REFERENCES

[1] Erik Duval's Weblog: "Learning Analytics and Educational Data Mining", https://erikduval.wordpress.com/2012/01/30/learning-analytics-and-educational-data-mining/ , 2012.

[2] "2016 Horizon Report", Educause Learning Initiative https://library.educause.edu/resources/2016/2/2016-horizon-report , 2016

[3] "Introducing Microsoft Power BI", Alberto Ferrari and Marco Russo, Microsoft Press, ISBN 9781509302284, 2016

[4] Business Ethics, Wikipedia article https://en.wikipedia.org/wiki/Business_ethics

[5] Tor (anonymity network), Wikipedia article https://en.wikipedia.org/wiki/Tor_(anonymity_network)

[6] "Privacy and Analytics – it's a DELICATE issue. A Checklist to establish trusted Learning Analytics.", Drachsler, H. and Greller, W., 6th Learning Analytics and Knowledge Conference 2016, April 25-29, 2016, Edinburgh, UK, http://dspace.ou.nl/handle/1820/6381

[7] Microsoft Online Services Privacy Statement, https://www.microsoft.com/EN-US/privacystatement/OnlineServices/Default.aspx, 2016

[8] Samsung Privacy Policy, https://account.samsung.com/membership/pp (translated from Croatian language), 2016

[9] Microsoft Education, https://www.microsoft.com/en-us/education

[10] "Microsoft to acquire LinkedIn", Microsoft News Center, ttps://news.microsoft.com/2016/06/13/microsoft-to-acquire-linkedin/, 2016

[11] "Privacy Fears Over Student Data Tracking Lead to InBloom's Shutdown", by Olga Kharif, Bloomberg http://www.bloomberg.com/news/articles/2014-05-01/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking, 2014

[12] "The Stanford education experiment could change higher learning forever", by Steven Leckart, Wired, https://www.wired.com/2012/03/ff_aiclass/, 2012

[13] Udacity, https://www.udacity.com/

[14] Udacity Nondegree Plus program https://www.udacity.com/nanodegree/plus, 2016