

Students' Attitudes toward Personal and Learning Data Usage in Aptitude Project Learner Taxonomy

Adelina Aleksieva-Petrova

Technical University of Sofia, Sofia, Bulgaria

aaleksieva@tu-sofia.bg

Abstract. Learning analytics (LA) is a process, which collect and analyze the learner data and activities in order to provide predictive indicators and increase the effectiveness of learning. The paper proposed the learner data taxonomy, which is used in Aptitude project to define main objects in learning analytics. Based on that taxonomy a survey is design and implement in order to study the students' attitude to using personal and learning data for LA. The results show that most of them would be provide information related with their academic background and learning experience information.

Keywords: Learning Analytics, Privacy, Learner Taxonomy.

1 Introduction

In each area, data analysis widely increased in order to make business decisions possible. This practice is most common in marketing, where users' behavior is analyzed in order to personalize their ads and provide appropriate recommendations.

Education is an area in which business influence is less prevalent, but nowadays more attention is focused to provide effective methods and technologies to achieve higher results in this area. Motivation for learning can be achieved by adapting and recommended appropriate learning contents and activities to learners. These processes could be supported by analysis of learning data, known as learning analytics (LA), which is provided by different learning systems and tools.

Learning analytics as statistical learning techniques are used to extract actionable insights from large data streams for optimizing teaching and learning [1]. With devices, systems and social media, a greater portion of the learning process generates digital trails, which offer an opportunity to explore learning from new and multiple angles [2].

This is also the main goal of the Aptitude project, studies and designs the platform for adaptation and recommendation of learning contents and activities

based on learning and gaming analytics. The innovative issues of the platform determine the main goals of the project as follows:

- Data studies – data acquisition regarding learning course modules from open-source learning management systems (LMS) such as Moodle and from smart adaptive educational games for the same learning course module, together with data preparation (cleaning) and storage for analysis purposes.
- Methodology definition – coining principles and procedures for the systematic pursuit of knowledge based on learning and gaming analyses of big data from LMS and educational games.
- Semantic modelling – construction of a formal semantic data model merging big data flows from LMS, together with an ontology for semantic recommendation and adaptations of both the learning content and workflow.
- Data analysis – run-time analysis of learning and gaming big data for providing descriptive, predictive and prescriptive results for an individual learning progress.
 - Learning courseware enhancement and platform development – adaptation and enhancement of both learning content and activity workflows.
- Validation of both the methodology and platform for big data learning and gaming analytics by practical experiments.

In order to achieve effective LA, on the one hand, a learner data taxonomy should be proposed. On the other side issues for LA fall into the following categories: the location and interpretation of data; informed consent and privacy of data; and the management and classification of data [3]. Therefore, the students' attitude for using of their personal and learning data for LA needs to be studied.

Thus defined the main propose of this study, namely students' attitude towards utilizing personal data and data generated as part of their learning activities for the goals of improving learning content and activities.

In order to achieve paper goal the methodology is proposed which follow the paper sections. The next section presents related works in two main point of view: LA as process, which need personal, and learner data and some privacy and data protection in LA. The third section proposes Aptitude learner data taxonomy using as sources and examples different systems and tools. In fourth section, survey for the students' attitude to using personal and learning data for LA is designed and presented. The fifth section summaries the results and discusses them. The last is conclusion.

2 Related works

Learning analytics have converged with educational data mining as increase the focus on student behaviors over the last five years [1]. LA divide into five

categories: learning, teaching, administration, technology development and digital citizenship. The last category, digital citizenship, affects ethical and privacy issues [4].

The gathering and use of students' data and their learning is providing new opportunities for institutions to support learners and to provide predictive indicators for attainment. Finally yet importantly learning analytics increases the quality and quantity of feedback loops in the education system for all participants in learning process [5].

One of the problems around learning analytics is the lack of clarity about what exactly should be measured to get an understanding of how learning is taking place. Some of the typical measurements include time spent, number of log-ins, number of mouse clicks, number of accessed resources, number of artifacts produced, number of assignments, etc. [6].

Despite the popularity of learning analytics, there remains significant barriers and challenges in organizational adoption [5]. Learning analytics functions that include student profiling entail a higher risk and attention must also be paid to the legal protection of students, which in the case of learning analytics means, above all, that the quality of the data must be ensured [7].

In [8] survey is provided to study the importance of using personal data for learning analytics. The results shows that personal data collection is most useful when used for the continuous improvement and personalization of the learning process. For example, demographic data analysis could determines the potential demand for education and also the nature and type of education to be provided [9].

The result of an explorations for privacy and data protection for LA are define the following principles:

- privacy and data protection in LA are achieved by negotiating data sharing with each student;
- openness and transparency are essential and should be an integral part of institutional policies; and
- big data will impact all society and to strengthen their personal data literacies [10].

Jones [11] proposes model of informed consent by improving the existing technical identity layer with Platform for Privacy Preferences technology and creating privacy dashboards that enable student to set privacy preferences and works to support student privacy and autonomy.

3 Aptitude learner taxonomy

In order to define main learner data the logs, databases and/or reports from the different systems, tools and web services are studied.

In class of Learning/course management systems, Moodle database and logs are explored. The number of tables in database is enormous and almost 10 percent of them refers to user information. The other tables play a key role in the system and contain information that is needed for courses and different types of activities such as forum, chat, assignments, glossary, book, wiki, etc. [12]. According to the process of system backups there are three main database tables related to it: backup_controllers, backup_logs, and backup_courses.

One of the main resources for personal data storage are the logs in Moodle, which are nothing more but tables, filled with actual student’s activities. Logs are available at site level and at course level and may have any combination of group, student, date, activity, actions and level [12].

In class of collaboration app and for videoconference meeting MS Teams is analyzed. MS Teams analytics gives reports in three levels: cross-team analytics, per-team analytics and per-channel analytics that are defined in concept of MS Teams system. In general, all these levels include the number of active users, posts (chat), replies, apps and/ or meeting in the specified period.

The proposed learner data taxonomy is presented in Fig. 1. The Learner Related Information is composed from two main classes: Personal Information (PI) and Academic Information (AI). Personal data can be categorized into two main groups: private information and demographics data. The private information is most sensitive data [13] but includes main identification data for learner as names, emails, student-identification numbers, social security numbers, some digital files such as photographs and other forms of information that may reveal a specific learner’s identity. The demographics data includes some additional information as an address, a date and place of birth, race, gender, economic status and others.

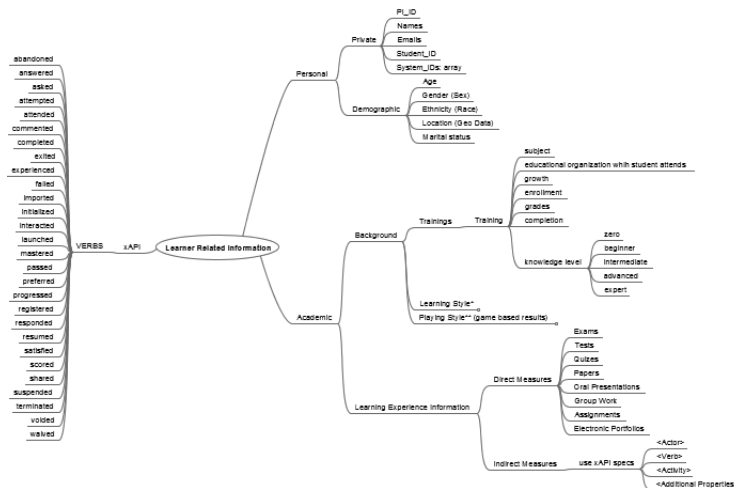


Fig. 1. Aptitude Learner Data Taxonomy.

The second class is Academic Information and includes data related with academic background (such as the educational organizations which a student attends, courses, enrollment, grades, completion, etc.) and various other forms of data collected for learning experience including evidence of learning outcomes (formal and informal) and learning activities (attendance, behavior, extracurricular activities, program participation, etc.).

In order to simplify the process of survey for the participants three top levels from proposed taxonomy are used for the survey. In survey, the rest of the taxonomy is applied as listed examples in question for indicated classes.

4 Survey for the attitude to using personal and learning data for LA

In order to studied, the students' attitude for using of their personal and learning data for LA the design of the survey is provided. The survey includes question with four answers and participants could choose multiple the items, which they agree with.

The question which look for opinion is “Which of your data would you agree to be used to adapt and recommend learning content and activities?”. The four possible answers are proposed to participants:

- personal information (e.g. name, email, address, age, gender, race, learner residence)
- academic background (e.g. educational institutions attended by the learner, current level of learner and years of attendance, levels of education, etc.)
- learning experience information (e.g. courses completed by the learner; course test and assignments grades and achievement; academic requirements completed by the learner; extracurricular activities, etc.)
- other learner data (e.g. information related to disciplinary problems, medical and health problems, etc.).

The target group of the survey is undergraduate students in bachelor degree program “Computer and Software Engineering” which use an online learning platform Moodle for learning content and assignments, MS Teams for videoconference meeting and YouTube for watching lectures.

The study involved 68 undergraduate students who received a questionnaire delivered electronically (by email) or by hand. They were asked to give their opinion on these four statements. If the students are agreed to provide some of the data their answer is note as 1, otherwise as 0. The questionnaires were collected and the results were summarized in electronic format.

5 Results and discussion

Fig. 2 shows a sieve diagram for the survey set and is visualized frequencies in a two-way contingency table and comparing them to expected frequencies under assumption of independence. The area of each rectangle is proportional to the expected frequency, while the observed frequency is shown by the number of squares in each rectangle.

The plot shows that the two variables, Academic background and Learning experience information, are highly associated, as there are substantial differences between observed and expected frequencies in all of the four quadrants.

The score combinations, which make the ranking of attributes, are shown in Fig 3.

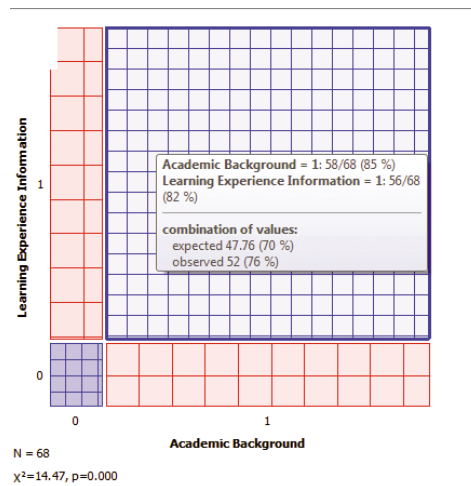


Fig. 2. Sieve diagram for the survey set.

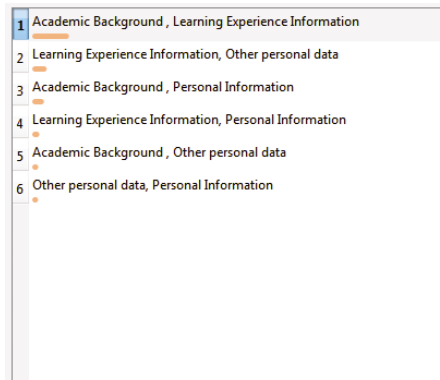


Fig. 3. Score combinations for the survey set.

The results of the survey show that 5.88% of the learners do not agree to provide any information and only 39.71% agree to provide and personal information in additional (see Fig. 4). The highest percentages are the students who agree to provide for usage of their academic background (85%) and learning experience information (82%).

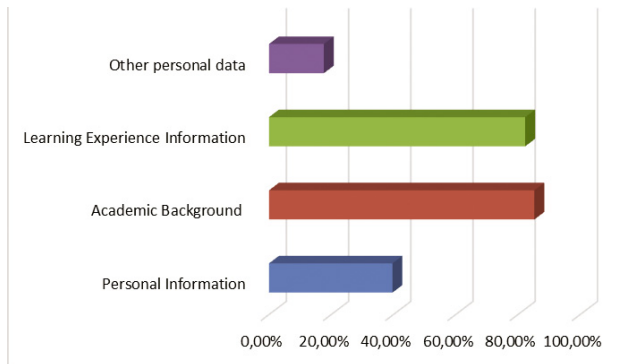


Fig. 4. Percentage of approve for the provision of separate data.

Fig. 5 shows different combinations of participants' answers and their percentages. Most of the students agree to academic background, learning experience information and other information except personal information (60%). Only 26% are those who want to provide personal information with academic background and learning experience information.

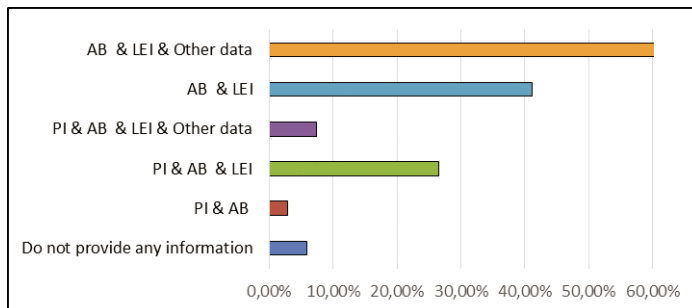


Fig. 5. Combinations of participants' answers and their percentages.

6 Conclusion

In the training of each person, it is necessary to offer the opportunity for adequate adaptation and recommendations of learning content and activities. One way to achieve this is by using LA.

One of the paper goal was to propose a taxonomy of the learners' data that should be collected and analyze in the process of LA. However, some of this data is private and sensitive. A study is conducted among 68 students, which shows that the vast majority of students agree to provide some form of data and with more than one third agreeing to include personal data.

The survey among students shows that a few of them are agreed to provide their private personal data (as names, emails and etc.) in order to receive adapted training with the possibility of recommendation. Most of them tend to provide information such as academic background and learning experience information to achieve this, but while maintaining their anonymity.

A taxonomy is developed for the purposes of the survey that helps classify the types of student data to be shared.

As future work, the paper results and proposed taxonomy will be used for LA. That will be implemented in Aptitude platform for recommendations and adaptation of learning contents and activities.

7 Acknowledgment

The research reported here was funded by the project “An innovative software platform for big data learning and gaming analytics for a user-centric adaptation of technology enhanced learning (APTITUDE)” – research projects on the societal challenges – 2018 by Bulgarian National Science Fund with contract №: KP-06OPR03/1 from 13.12.2018.

References

1. Lemay David J., Baek C., and Doleck T. “Comparison of Learning Analytics and Educational Data Mining: A Topic Modeling Approach.” *Computers and Education: Artificial Intelligence* (2021): 100016.
2. Siemens G. “Learning analytics: The emergence of a discipline.” *American Behavioral Scientist* 57.10 (2013): 1380-1400.
3. Amara A., et al. “Learning analytics in higher education: a summary of tools and approaches.” *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference. Australasian Society for Computers in Learning in Tertiary Education*, 2013.
4. Zhong L. “A systematic overview of learning analytics in higher education.” *Journal of Educational Technology Development and Exchange (JETDE)* 8.2 (2015): 3.
5. Joksimović S., Kovanović V., and Dawson S. “The journey of learning analytics.” *HERDSA Review of Higher Education* 6 (2019): 27-63.
6. Duval E. “Attention please! Learning analytics for visualization and recommendation.” *Proceedings of the 1st international conference on learning analytics and knowledge*. 2011.
7. Ouli J., and Voutilainen T. (2019), “Learning analytics and the processing of student data in universities”, *Edilex* 2019/37
8. Aleksieva-Petrova A., and Petrov M. “Survey on the importance of using personal data for learning analytics and of data privacy.” *2020 International Conference Automatics and Informatics (ICAI)*. IEEE, 2020.

9. Obasi Kenneth K. "Demographic Data Analysis And Educational System Planning For Primary Education Delivery In Abia State, Nigeria." *Advances in Social Sciences Research Journal* 6.2 (2019).
10. Hoel T., and Weiqin C. "Privacy and data protection in learning analytics should be motivated by an educational maxim—towards a proposal." *Research and practice in technology enhanced learning* 13.1 (2018): 1-14.
11. Jones K.M.L. "Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy." *International Journal of Educational Technology in Higher Education* 16.1 (2019): 1-22.
12. Aleksieva-Petrova A., I. Chenchev, and M. Petrov. "LMS Data Collection, Processing and Compliance with EU GDPR." *EDULEARN19*. 2019.
13. Aleksieva-Petrova, A., I. Chenchev, and Petrov M. (2020), "Three-Layer Model for Learner Data Anonymization", *Proceedings of 14th International Technology, Education and Development Conference*, url: <doi: <http://dx.doi.org/10.21125/inted.2020.1688>>.