Chapter 7
The LAPS Project: Using Machine Learning Techniques for Early Student Support

Mathias Hinkelmann and Tobias Jordine

1 Introduction

Early support of students can be a successful instrument to improve the academic achievements. Universities are challenged to identify students who may take advantage of early support from the university's student and learning support center. This chapter will introduce the LAPS project ("Learning Analytics für Prüfungsleistungen und Studienerfolg"/"Learning analytics for exams and study success"), developed and used at the Hochschule der Medien (HdM), Stuttgart, Germany, which is set up to cover these challenges. The particular approach of LAPS project is that completed study progressions are analyzed via machine learning techniques. These results are compared to the grades reached by the students in their study program so far. Since the progression of an enrolled student will statistically not differ from students, who either completed or failed in their study program, the comparison can be used to make an individual statement about students' risk of failure or possibility of success. Based on the findings of the risk calculation, students can be advised more focused. In addition, the findings support both, under and top performing students. Also, the results of the statistical analysis by the LAPS software can be used as a factual basis for discussions aiming at improvements of the study programs. This chapter is structured as follows: The Sect. 2 introduces current approaches and projects in the learning analytics research area. Section 3 provides insights into several aspects of the presented project: the used data basis is shown, technical implementation details are provided, and findings of the feasibility study are presented. Further, the privacy and ethical considerations that were identified are explained. The section concludes how LAPS can be used for academic quality

M. Hinkelmann (✉) · T. Jordine
Department of Computer Science and Media, Hochschule der Medien, Stuttgart, Germany
e-mail: hinkelmann@hdm-stuttgart.de

assurance. The software and the LAPS process are reviewed in Sect. 4. Finally, Sect. 5 concludes this chapter by providing information about possible improvements concerning the LAPS project.

2 Existing Work

Based on Ferguson (2012), the research area of learning analytics has its roots in business intelligence, web analytics, educational data mining, and recommender systems and is defined as follows:

> Learning analytics is 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. ("1st International Conference on Learning Analytics and Knowledge 2011," 2010)

Besides technological aspects like data collection and big data analysis, learning analytics must be seen as holistic as technology, socialization, and pedagogy are involved (Siemens, 2010). Figure 7.1 shows the process of a traditional learning analytics approach.

As the LAPS project is aiming at the German academic system, related German approaches are presented below.

- *Study progression analysis approaches*
  In the case of study progression analysis, the focus is on comparing the progression of individual students with the study plan and the progress of the entire group of students in a study program. This approach, which is pursued at many universities, is represented by the "tempo 30" project of the Ravensburg-Weingarten University or "StuVa" at the University of Freiburg (Hermann & Ottmann, 2006). A special approach to study progression analysis is module-based monitoring (Jaeger & Sanders, 2009). In this particular approach, budget-oriented views on university management and approaches to ensuring the quality of teaching are considered.
- *Predictor models*
  With this approach, a presumed predictor for academic success is analyzed in detail. At the Kiel University of Applied Sciences, for example, two study programs were
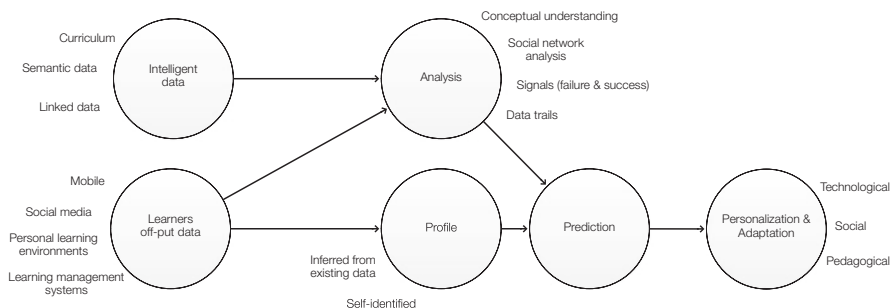

Fig. 7.1 The learning analytics process (Siemens, 2010)

used to examine the success of studies in the first semester as a predictor for overall success and to show that early indicators are indeed present. These indicators can be used to control advisory services (Christensen & Meier, 2014). Other studies (e.g., Trapmann, Hell, Weigand, & Schuler, 2007) have investigated the extent to which school grades can be used as predictors. But recent developments in the field of increasingly heterogeneous access to higher education show that simple predictors are no longer sufficient. Instead, multidimensional predictors must be used.

3 The Laps Project

3.1 Data Basis

Efficient administration of universities requires the use of Campus-Management-Systems (CMS) which allows to track and support the entire Student-Life-Cycle starting from his/her application till his/her de-registration. One of the key elements of such a CMS is the management of exams. Because of the required legal certainty, all CMS record all exam-related student data. This means that a high data quality in terms of students' master data as well as collected exam data is available. Additionally, retention periods given by legal provisions lead to a large data basis.

The development of this data basis adds value for the organization of study and exam regulations. For doing so, personal data like type and grade of the university entrance qualification, date of enrollment, date of de-registration, as well as detailed information about students' exams can be used and analyzed. Due to the similarity of the tasks and requirements for a CMS, it can be assumed that the considerations for the indexing of data within the CMS are not limited to specific systems (e.g., the products of HIS eG[1]) but are directly transferable to other CMS. Table 7.1 shows typical data available in Campus-Management-Systems.

Table 7.1 Typical data available in campus-management-systems

Student-related data	Exam-related data
• Gender • Birthday • Grade of the university entrance qualification • Type of the university entrance qualification • Date of the university entrance qualification • Course of studies • Date of enrollment • Date of de-registration • Study success	Data on academic performance • Name of the academic achievement • Number of ECTSs awarded • Assignment to a study stage • Mandatory/optional/additional academic achievement Data on the trail • Date of exam • Assignment to a semester • Exam result • Attempt counter • Identification of withdrawals from exams

[1] https://www.his.de/.

3.2   The LAPS Approach

An approach that uses the data described in the above is developed at the Stuttgart
Media University since 2014. Analyses that were made in advance of the develop-
ment of LAPS have shown that:

1. A combination of the type of higher education entrance qualification, the grade
   of the higher education entrance qualification, and the time interval for admis-
   sion to a course of study.
2. The gender of the students has a measurable influence on the probability of drop-
   out. Analysis performed at HdM in advance of the LAPS project has shown that
   male students have a higher risk to fail. This finding is independent from the
   percentage of male and female students in a study program.

Overall, preliminary studies have shown that simple, experience-based predictors
are not sufficient to identify critical study situations (Trapmann et al., 2007) and that
a systematic and multidimensional analysis of the data (students' master data and
data on the examination events) is required. This requires an automated, algorithmic
evaluation. The LAPS software therefore uses machine learning methods. During
machine learning, patterns are not set manually but are "learned" automatically
from existing training data.

The transfer of this approach to the analysis of study situations is possible due to
the existence of completed study progressions. Data of de-registered students are
used to determine specific study situations. This approach is explained in the fol-
lowing, starting from the general principle of an automated learning process illus-
trated in Fig. 7.2.

In the training phase, a model is trained which is used to calculate forecasts and
classifications. As training data, the LAPS software uses the enrollment data, study
progress data, and study success data of all students who have already completed
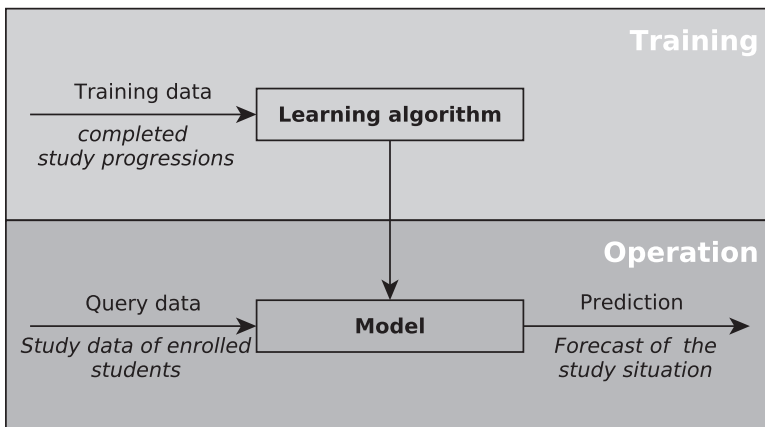

Fig. 7.2 The principle of machine learning

their bachelor studies. Each of these students is described by a data record, which in turn consists of a list of characteristic feature/value pairs: For example, "gender = female and type of university entrance qualification = Abitur" are two feature/value pairs of the enrollment data and "not successful exams after the first semester = 1 and ECTS after the first semester = 20" are two feature/value pairs of the study progression data. The machine learning method used in LAPS is the Apriori algorithm (Agrawal, Imieliński, & Swami, 1993). The model calculated by this algorithm is a set of association rules. Each association rule describes a frequently occurring combination of characteristic feature/value pairs in the form of an implication: A → B. In general, both premise A and conclusion B can represent any subjunctive link between characteristic feature/value pairs. A possible rule would be, e.g., "(number of exams graded with fail after 2nd semester = 3 and number of ECTS after 2nd semester < 20) → studies successful = no."

A rule is only recognized as relevant by the learning algorithm and included in the model if its support and its confidence are greater than a minimum value that can be set by the user. The support describes the composite probability P (A, B), i.e., the relative frequency with which the premise and conclusion occur together in a training data set. On the other hand, confidence describes the conditional probability P (B|A), i.e., the relative frequency for which the conclusion is also true in training data in which the premise is true. Support is therefore a measure of whether the pattern consisting of A and B occurs frequently enough in the training data to be considered statistically relevant. The confidence specifies the certainty with which rule A → B applies.

All association rules whose support and confidence are greater than the minimum values form the trained model. After training, the trained model is used as follows to predict a critical course of study: For a student to be analyzed, all currently available characteristic value pairs are entered in the system as a query. For this specific query, the subset M_S of the association rule set contained in the model is determined, for which the premise A is fulfilled with the entered student data. Since the conclusion of all the rules contained in the model is constantly "study successful = no," the confidence of each rule in M_S indicates the probability with which the respective rule predicts an unsuccessful completion. The median is calculated from the confidence values of all rules applicable to the student (quantity of M_S). This median represents a preliminary risk probability for the student. After the preliminary risk probabilities have been calculated for all students, they are adjusted for the purpose of better differentiation by assigning the final risk score of 100% to the student with the highest risk and 0% to the student with the lowest preliminary risk score. The final risk values of all other students are derived from their preliminary risk values by linear scaling.

By using the Apriori algorithm, a large number of possible risk dimensions with different characteristics can be defined for the analyses in LAPS. In this definition, it is not required to consider the relevance of the analysis dimension or characteristics. This task is performed during the training phase, in which the relevance of combinations of these characteristics is determined. The following list shows the currently used risk dimensions used by the LAPS software:

Dimensions of personal data:

- Age at beginning of study
- Gender

Dimensions of educational biography:

- Type of the university entrance qualification
- Date of the university entrance qualification
- Time interval between acquisition of the university entrance qualification and start of studies

Study course analysis dimensions by semester:

- Total of achieved ECTS points
- Achieved average grade
- Number of failed exams
- Number of successful exams
- Number of deferred exams
- Frequency of nonappearances in enrolled exams

The course of study dimensions grouped by semesters is additionally assigned different characteristics. The risk dimension sum of the achieved ECTS credits is analyzed after the first semester of studies with the following characteristics:

- <10 ECTS credits
- <20 ECTS credits
- <30 ECTS credits
- More than 30 ECTS credits

This results in more than 200 possible individual risk characteristics, which are linked in the training phase on the basis of completed courses of study and lead to the analysis model, which comprises several thousand combinations of risk dimensions. In the current version of LAPS, the risk dimensions and characteristics can be configured. It is the responsibility of user of the system to define a threshold for the predicted risk value above which the affected students are automatically classified as critical. In LAPS, student and examination data is updated every semester via a file upload interface incrementally. After the import, a training phase takes place automatically, which is followed by the analysis of the currently enrolled students. The described recognition of critical study progressions by predicting the risk of termination does not represent the only application of the model trained in LAPS. The learned rules are also able to identify typical patterns of under- and overstraining or frequent postponements of examinations.

Besides the functionality to identify risks, LAPS is also capable to identify study progressions with a high potential. This information can be used to support top performing students, e.g., with a fellowship or additional classes. To identify these students, achieved ECTS credits per semester are calculated and compared with set point of ECTS credits. If the achieved ECTS and the current average grade are significantly better than the mean of the cohort, respectively, the set point of ECTS credits, the student is identified as a top performer.

3.3   Feasibility Study and Use in Consultations with Students

At an earlier stage of the project, students' risk data were analyzed and used in consultation situations by staff members of the student support center and course leaders. In contrast to the current version of LAPS, students were directly contacted by the users of the system when a risky study progression was identified. This version of the tool only supported the identification of risks and was not able to detect positive study progressions.

Having this setup, students were invited for a consultation discussion. The results of the risk analysis served as an evidence-based foundation of this talk and helped students to understand their situation. This was especially useful when students had a different impression on their study progression. It was found out that by using the LAPS software, students can be advised at an earlier stage of their studies and can be one addition to reduce students' dropouts as additional support like trainings or adjustments of the study progression can be offered.

Users' feedback of this early version of LAPS was positive. It was liked that in contrast to traditional grade overviews, the LAPS profiles are much more detailed and potential risks are immediately visible. This allows to develop individual counteractions. But the feasibility study also showed that the handling with students' personal data was not ideal, since lecturers (i.e., persons who do the grading as well) can access and view students' risk details without their permission. This is why it was required to define premises for the privacy and ethics for the project, which will be explained in detail in the following section.

3.4   Privacy and Ethics

The LAPS software serves to create an evidence-based discussion basis with students at an early stage of their studies. This evidence-based approach contrasts with legitimate data privacy aspects. For the LAPS project, privacy and ethical premises are a foundation of the whole project. These premises are voluntariness, self-determination and self-responsibility, respecting individuality, confidentiality, as well as anonymity and are taken into account in several ways, which are explained in the following. Figure 7.3 provides an overview of the LAPS data access process.

When students de-register for any reason, their personal data is no longer visible for any user of the system. In the case of enrolled students, students must opt-in to be considered by the system. Only with their explicit agreement it is possible to view their personal data and risk analysis. Students always have the chance to change their decision whether they take part or not. New students can take part at the LAPS project during enrollment, whereas current students are informed via e-mail. Transparency is very important as students get informed how their data is used. For example, in advance of the opt-in, a privacy information sheet that explains the use of data is presented to each student. Additionally, the project is presented at the general student meeting each semester as well as an information booth where students can ask project members about LAPS once a semester.
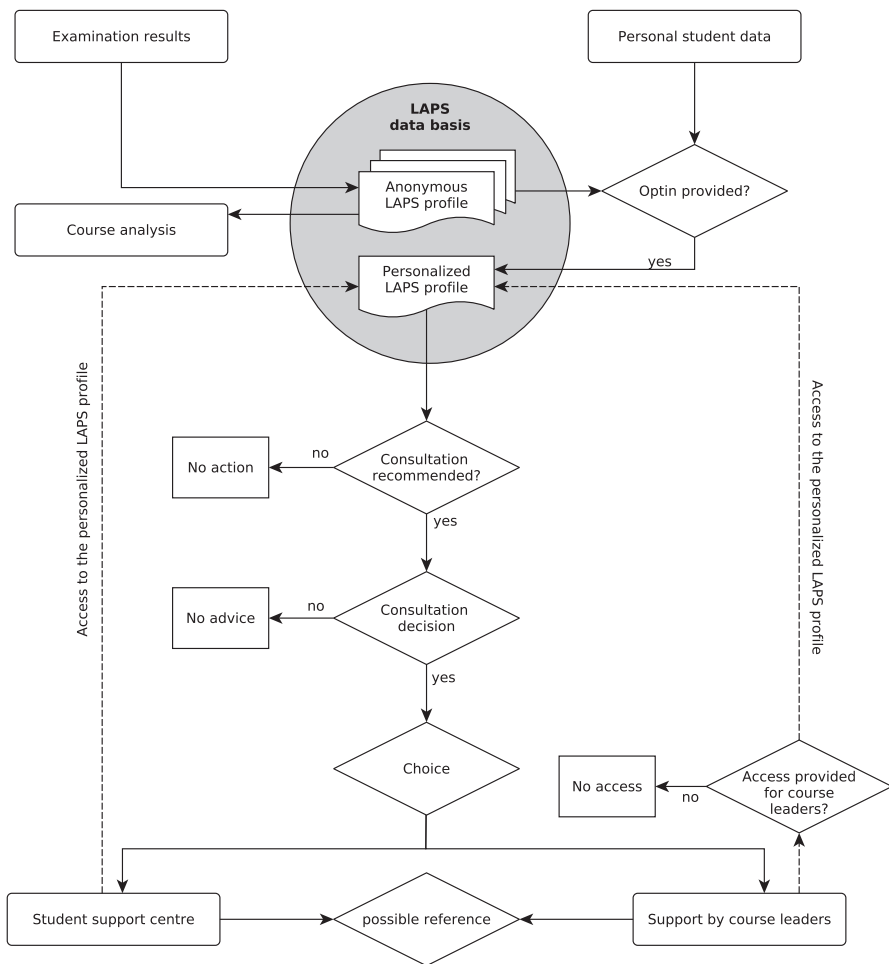
Fig. 7.3 The LAPS process

Access to the data is strictly limited by the limitation of the user group. In the case of the HdM, course leaders have access to the personal student data of their respective course after they have taken part at a LAPS consultation introduction workshop. This workshop is aiming to help course leaders to understand the data and analysis results calculated by the system and how they can use this information for a successful consultation. Besides course leaders, staff members of the student support center have access to the results of the students who agreed to take part at the LAPS project.

When a risky or an exceptional good study progression is identified, students are informed via an automatically generated e-mail. After receiving the e-mail, the decision is up to the students to ignore it or to choose an individual consultation discussion with either members of the student support center or their according course leaders. As part of the ethical and privacy decisions of LAPS, students will

not directly get the results of the analysis. This is intended to prevent self-fulfilling prophecy: Without having the knowledge how to interpret the analysis and identifying specific needs and students' personal life situation, the results could be misunderstood as the algorithm is only able to do calculations based on data stored in the Campus-Management-System.

The project is already compatible with the EU-DSGVO (General Data Protection Regulation, 2018). This ensures that the project complies with the currently valid data privacy laws.

3.5   Functionalities of the Tool to Support Students

The relevance of the data on the enrolled students increases with the import of the examination results from the previous semester. At the HdM, the system is updated in the seventh week of the lecture period of the following semester onward due to administrative constraints. After the update of student data, a list of critical study progressions of participating students can be reviewed. The displayed data is initially anonymized in the list view (all students) as well as in the individual view (individual student). The de-anonymization of individual cases must be done consciously by clicking a button and is only possible when the student takes part at the project. This should limit bias effects with regard to the identity of the individual student. This detailed view provides the advisor compact information about the student (see Table 7.2).

Table 7.2 Detailed student information in LAPS

General	• Course of studies • Enrollment date and type of the university entrance qualification • Start semester • State of studies
Student	• Birthday • Age at enrollment • Gender • Name (after explicit de-anonymization) • Banner ID • E-mail address • Type of the university entrance qualification • Grade of the university entrance qualification • Date of the university entrance qualification
Examination data	• Number of examinations • Successful examinations • Failed examinations • Excused cancellations • Total ECTS • Current average grade • Average grade base studies • Average grade main studies • Risk score

76.0% (13.3%) 861 / 1133 / 6485	Semester with examinations >=1	Type of university entrance qualification **College**	Not successful in first semester >1	Successful study? **No**

76.0%: **(Confidence)**: Failure probability
13.3%: (Support): Frequency of a case of this constellation (filter criteria & risk criterion)
861:    Number of students to whom the filter criteria and the risk criterion apply
1133:   Number of students to whom only the filter criteria apply
6485:   Number of students considered

Fig. 7.4 Identified risk and its representation in LAPS

This information is supplemented by a report and presentation of the actual status of the examination results and the course of studies at various levels:

- The semester table lists the acquired ECTS, the ECTS total, the average grade (weighted according to ECTS), and the number and status (passed, failed, approved cancelled) of the examinations taken per semester.
- An overview of the registered examination performances of the previous semesters is provided.
- The performance chart lists detailed information on all examination (e.g., ident number, description, status, ECTS, grade). By clicking on the ident number of an examination, a detailed view is presented, and the grades for all available semesters can be seen. In this way, the student's performance can be compared to the overall cohort.

The individual view is completed by the risk details: a graphical representation of the distribution of risks (with which frequency risk criteria of a certain probability of failure apply to the student) as well as a representation of the risk criteria applicable to the student. Figure 7.4 provides an example of an automatically identified risk.

3.6 Using LAPS for Quality Assurance

Besides the functionality to support students based on the LAPS profile, the software supports quality assurance of study programs. The following functionalities are designed to provide information about specific programs, lectures, and student cohorts. The analysis results presented are based on the anonymous LAPS profiles, which mean that personal data is not visible.

- *Programs*
  In this view, study program information can be obtained. The following data is available for each program: number of enrolled students, number of dropouts, number of successful study progressions, average risk possibility, minimum/average/maximum student age, gender distribution, average grade of the university entrance qualification, and retreats from examinations.

- *Cohorts*
  For the development of study programs, it is important to get information on the consequences of the changes on module level, e.g., to the examination regulations and the curriculum. The cohort's view allows to compare the distribution of students obtained ECTS credits per semester and to identify possible structural problems when students do not achieve the required ECTS.
- *Lectures*
  This view allows an in-detail analysis for each semester of lectures and provides access to distribution of grades, number of successful examinations, average grade, number of retreats, and number of registrations.

4   Discussion

Although the LAPS project is developed at HdM, it is open to be used at any other university. This achieved by being released as an open-source software and the data import is not bound to a specific CMS. The only requirement is that the CMS data needs to be exported into a LAPS-readable CSV format. For doing so, it is required to write export scripts that allow to export the data. Additionally, it could be possible to adjust the definitions mentioned in the above as the study progression differs from each university.

Nevertheless, some lacks and points of discussion were identified for the project, which are described in the following:

- *Using students' gender as part of the risk calculation.*
  As a part of the risk analysis, the LAPS software uses students' gender information. It is not intended to make differences or judgements between the genders. In fact, the risk analyzation results can identify potential problems of gender groups.
- *Validity of the used data model.*
  When the LAPS software is used for consulting students, advisors need to be clear about the underlying data that are used to calculate the possibility of a positive/negative study progression. As the algorithm considers sociodemographic and examination data, all derived risk probabilities are based on these facts. During a consultation situation, it is required to know that the algorithm may identify a student progression as risky which is due to a small number of ECTS credits obtained during the first semesters. This could have multiple reasons, e.g., illness of the student. To cover this issue, the LAPS software is inextricably bound to the LAPS consultation process which includes a mandatory consultation introduction workshop.
  Illnesses that result in a long-lasting absence from university, such as illnesses that are part of the general risk of life, such as influenza, etc., usually result in vacation semesters and are considered accordingly in the LAPS risk assessment. The underlying data for such events is recorded by the CMS. Other data, e.g.,

other health issues and possible labor of students, is not recorded due to the strict privacy laws in Germany. These students' information can be taken into consideration during the consultation, which is part of the LAPS project.

- *Low response rate for course leaders taking part in the LAPS consultation introduction workshop.*
  As described in the above, course leaders need to take part in mandatory workshops that provide an introduction to consultation of students using LAPS data. All course leaders of the bachelor programs were invited, but only a few of them responded. The main problem was that the workshop was planned during semester holidays and many of the invited course leaders were not available for two consecutive days. This could be improved by splitting up the workshop into smaller lessons (e.g., 3 × 3 h) during semester.
- *Student response rate could have been better for the first run.*
  For the first run of LAPS, all students were invited via e-mail. For the targeted student group (all students in the fourth semester or below, $N = 1500$), 98 students (6.53%) participated. It is planned to increase the number of participating students by integrating the opt-in registration into the enrollment process. Nevertheless, only a single student of the students who filled the opt-in form did not want to participate.
- *Data analysis is CPU intensive and requires time.*
  As the analysis is complex data mining process, the calculation should be integrated into a batch job executed when the application is not used interactively, e.g., during nighttime.

5   Future Work

For the further development of the LAPS project, it is planned to analyze data on how students accept the system. It has to be tracked how many of the students who choose the opt-in are receiving a system-generated e-mail. Last but not least, it has to be tracked how many of them are taking advantage of the conversation offer with either staff members of the student support center or their according course leaders. A consultation guideline is currently developed to give support on how to use the system in such situations.

Technically, an automated ETL process (extraction, transformation, load), which is a standard process model for data warehouse and big data computing, could improve the upload of new student and examination data. By adding a functionality to track student progression within a lecture (e.g., by integrating the results of interim tests), students would be able to get information about their progression in a specific lecture. In addition to automatically identified risks, a manual student tagging functionality could extend the LAPS software: student progressions are anonymously presented to course leaders, and based on their experience, they could decide if the student needs additional support.

References

1st International Conference on Learning Analytics and Knowledge 2011. (2010, July 22). Retrieved February 26, 2018, from https://tekri.athabascau.ca/analytics/

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *ACM sigmod record* (Vol. 22, pp. 207–216). New York: ACM.

Christensen, B., & Meier, J.-H. (2014). Zur Frühidentifikation von Studienabbrüchen. *Das Hochschulwesen, 6*, 182–185.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304. https://doi.org/10.1504/IJTEL.2012.051816

General Data Protection Regulation (2018). (EU) 2016/679 §. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=DE

Hermann, C., & Ottmann, T. (2006). StuVa–Werkzeugunterstützte Studienverlaufsanalyse zur Unterstützung der Studienberatung. In *HDI* (pp. 127–136).

Jaeger, M., & Sanders, S. (2009). Kreditpunkte als Kennzahl für die Hochschulfinanzierung. In *Hannover: HIS–Forum Hochschule*.

Siemens, G. (2010, August 25). What are learning analytics? Retrieved February 25, 2018, from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Trapmann, S., Hell, B., Weigand, S., & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie, 21*(1), 11–27. https://doi.org/10.1024/1010-0652.21.1.11

---

[2] http://s-beat.de.