Analyzing Learning and Working Behaviour of Students with E-Learning Support

Individual Project
Project Report

presented by
Larissa Haas
1417669

supervised by and with data from
Daniel Klasen
Chair of Economic and Business Education - Learning, Design and Technology
University of Mannheim

submitted to
Prof. Dr. Heiko Paulheim
Chair for Data Science
University of Mannheim

September 7, 2018

Abstract

With proceeding digitization in educational realms, today's learning more and more involves the use of Learning Management Systems (LMS). These systems (like ILIAS in the actual case) can be customized to individual needs and accessed and used with an ordinary browser. Within this online e-learning system, lecturers, tutors, and students are able to communicate and work together in various different ways. All this actions and connections between users, as they are happening online, are recorded and stored and can be therefore used for data analysis.

The underlying goal of this project and the related report at hand is to connect the data of LMS usage of students to their exam results, and to answer some of the following questions en route:

- How do students work with the e-learning system?
- Do more active online-users yield an advantage for their studies?
- Can we predict the failure of students (and how early during the ongoing semester)?
- What are the possibilities of these methods, given we had more data?

This project took place in the Spring Semester 2018 under the supervision of Prof. Dr. Heiko Paulheim (Chair for Data Science), and in close collaboration with the Computing Center and Dr. Daniel Klasen (Chair of Economic and Business Education - Learning, Design and Technology), who provided access to the collected data.

Larissa Haas (1417669) - larhaas@mail.uni-mannheim.de

Contents

Larissa Haas (1417669) - larhaas@mail.uni-mannheim.de
Individual Project - Spring Semester 2018 - September 7, 2018

Chapter 1

Introduction

With proceeding digitization in educational realms, today's learning more and more involves the usage of Learning Management Systems (LMS). These systems (like ILIAS in the actual case) can be customized to individual needs and accessed and used with an ordinary browser. Within this online e-learning system, lecturers, tutors, and students are able to communicate and work together in various different ways. All this actions and connections between users, as they are happening online, are recorded and stored and can be therefore used for data analysis.

In previously published papers we can observe the analysis of a broad variety of teaching and learning methods, while everything fits under the umbrella term of Educational Data Mining (EDM). With its ten years of existence[1] EDM is a rather young but flourishing field of research. More and more papers are published concerning the usage of online libraries, the selection of courses in universities or the discovery of student learner types. This trend may take place because Data Mining, as one of EDM's origins, is currently a widely recognized buzzword in general, in science as well as in society, and educational data is more and more available because of the widespread use of LMSs. As "higher education cannot afford not to use the data", the interest in this topic is huge (Pardo, 2014, p. 34).

Because the results of EDM are interesting on the one hand for lecturers who want to understand their students in their learning process and support them within the wider context of Learning Analytics.[2] On the other hand, there are psychologists and developers who want to understand how learning works in general and how individual learning and working behavior can be supported best by the provided tools and systems.

[1] First Educational Data Mining conference in 2008 (http://educationaldatamining.org/conferences/) and Journal of Educational Data Mining since 2009 (http://jedm.educationaldatamining.org/index.php/JEDM)

[2] Learning Analytics (LA) in comparison to EDM uses the data with a "holistic focus" together with a human-centred judgement and intervention (Baker and Inventado, 2014, p. 62).

Livieris et al. describe EDM as "an academic field of research that seeks to obtain, from the data stored in educational environments, new and useful information in order to develop and strengthen the cognitive theories of teaching and learning" (Livieris et al., 2018, p. 269). As this definition is rather wide than narrow we find **various methods and approaches** summarized within this term: visualization, classification, and prediction as well as clustering of student types (Baker and Yacef, 2009, p. 4).

We can distinguish between different **levels of learning** such as analysis of degree proceedings (Asif et al., 2017) or course progress (Costa et al., 2017), or between the **sources of data**, for example, derived from course selection, books or media usage (Masip et al., 2011), and working behavior (Sheard, 2011a). Also, different **kinds of courses** were under investigation: online (Lee, 2018), distance or blended courses (Costa et al., 2017), as well as synchronous or asynchronous (Tang et al., 2018) course models. Additionally, we observe different positions when it comes to **data from other sources**, supplementary to our "main" data. Depending on the case, existing literature uses on the one hand explicitly a lot of additional data about the students under investigation (Costa et al., 2017), on the other hand, no additional personal data at all to avoid the trace of prejudice classification and wrong causation expectations because of "using group-risk statistics" (Scholes, 2016, p. 941). This "marks only" approach also "enable[s] the university administration to develop an educational policy that is simpler to implement" (on degree level) (Asif et al., 2017, p. 178). Because of the broad range of used learning and teaching models in the literature, it is hard to find comparable instances to ours and the data we have at hand. But this makes us think of other alternative approaches, which may be interesting as well.

In our case, we investigated the course "Praktische Informatik I". The data we had at hand came from ILIAS as well as from the homework and exam results. The course was a so-called "blended" course, which means that there were physical meetings of the students, tutors, and lecturers, accompanied by a range of online tools such as file storage, quizzes, videos, and questionnaires. The students had to complete the homework tasks at a specific point in time, so we observed a more or less synchronous course, with the exception that students only needed to score half of the points in all assignments taken together. This opened the possibility for them to leave assignments out when they thought the remaining points would be enough to be admitted to the exam. In the following chapter, I will now briefly discuss how the data looked like and what additional insights I could draw from the variables at hand.

Chapter 2

Data

2.1 Online Interactions Measured

The course, which will be analyzed during this project, it is composed of a lecture and several tutorials. Each student was assigned to one tutorial he or she had to attend weekly. In this meetings, each student had to successfully complete a row of assignments to get accepted for the final exam. These assignments were done in teams of two, whereas the final exam was done individually.

The data from ILIAS was collected by modifying this online environment and tracking the behavior of students. Students attending the course "Praktische Informatik I" were asked to activate the tracking in ILIAS (if they wanted to). "Anonymous" was the default option, roughly 11% changed their settings to "pseudonymous". In the end, we resulted in 18 active students, their identity was hidden by a number, connected to their account by a hash function. Due to privacy protection issues, the hash values are designed as there cannot be any clue on the "real" student behind the data. Nevertheless, the hash values are useful to connect the different data sets across sources (these connections can be seen in figure 2.1). The remaining 218 non-active students all got the same identification number, so their actions were not distinguishable by a person. There were only 4 students who declined to be tracked in ILIAS at all. As it was possible to change the selection during the semester, these numbers only show the distribution at the end of the semester, when the student data was stored. During the semester we observe events from 26 different student IDs. The table below shows the variables at hand and the numbers of available pseudonymous students.

	lecture	**tutorial**	**variables**
student data	241 students	221 students	hash value, last update on, leap ID, number of events, tracking status
event data	33873 events	70614 events	student ID, resource ID, type, timestamp, content
resource data	112 resources	323 resources	description, leap ID, name, number of events, type, uri

Table 2.1: Overview ILIAS Data

Not all variables listed above were filled for each record. The sparseness, especially for name and description of the resources, was very high. If this would be filled out more widely, this would open possibilities, for example for a content analysis.

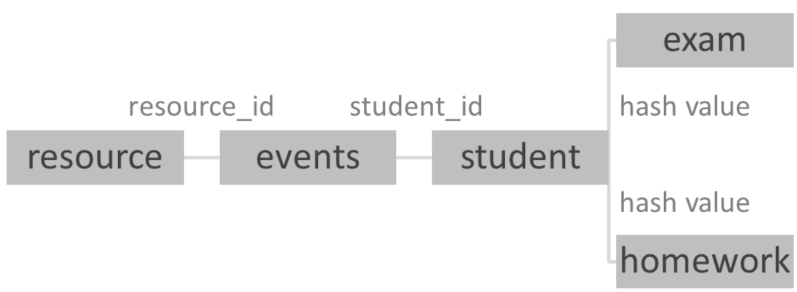

Figure 2.1: Datasets and connecting Variables

With the datasets "events" and "resource" I was able to create some other interesting variables, that may be helpful in later analysis. These new variables included the time between the create date of the resource and the event data (How long did it take after uploading something until some interaction happened?), as well as the variables weekday, hour, week number (extracted from the timestamp). I also added a status for students to be able to separate the pseudonymous students more easily. For the last step of our analysis, I condensed the event records for each of the 26 available non-anonymous students to one record (I preserved mean and standard deviation).

2.2 Course and Exam Results Measured

As seen in figure 2.1 ILIAS data and course data were connectable via hash values. Besides these values, the data offered the following variables, as seen in table 2.2.

		variables
homeworks	235 students	tutor, team ID, homework results 1-11, hash value
exam	218 students	study program, hash value, points from tasks 1-7, total points, final exam grade

Table 2.2: Overview Course and Exam Data

The information of variable "team ID" was created by hand, by comparison of the homework results (when within a team, the members scored equally). In addition, the team members were randomly sorted into groups 0 and 1 in order to compare the team members later. Homework number 11 was not mandatory, it was a bonus assignment for all those who were missing points to be admitted for the exam. The students needed at least half of the points from the homeworks (in total, not in each homework) to be admitted.

From the study program variable in the exam data, it was possible to extract information about the semester of the students, their anticipated degree, their subject and the faculty they were organized in. As degree, subject, faculty, as well as tutor, were text strings, they were converted into dummy variables to facilitate later analysis. Having a last glance at the course data, the results of the exam are rather expectable: Of 140 students with an exam grade available, the average grade was about 3.3 (ranging from 5.0 to 1.0), and roughly a quarter failed to pass the exam.

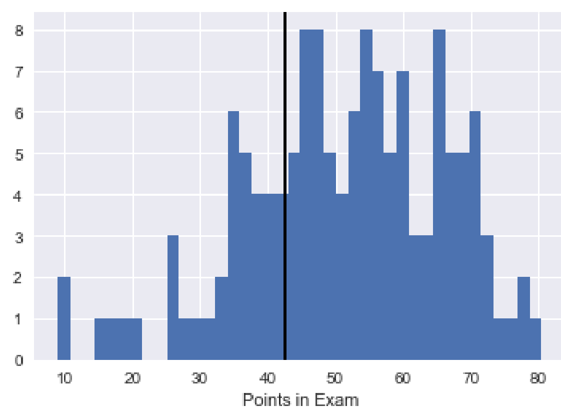

Figure 2.2: Frequencies of Points reached in the Exam

Chapter 3

Results

First, there will be a descriptive analysis of the data at hand and a comparison between anonymous and pseudonymous students. **Second**, the learning groups will be under investigation. And **third**, the impact of attitude to work and the usage of ILIAS through the semester will be analyzed with different data mining algorithms. Though other papers have shown (Costa et al., 2017) that Support Vector Machines result in best possible predictions of failing students for their student interaction data, I will not use them for this project. SVM models are widely considered as black box (Hämäläinen and Vinni, 2011, p. 69) and cannot be interpreted intuitively by humans. According to (Wanli et al., 2015, p. 169) this characteristic makes SVMs and other black box algorithms unsuitable for EDM:

> Model interpretability in performance prediction is important for two primary reasons [...]: first, the constructed model is usually assumed to support decisions made by human users - in our context, to facilitate model is a black-box, which renders predictions without explanation or justification, people or teachers may not have confidence in it. Second, if the model is not understandable, users may not be able to validate it. This hinders the interactive aspect of knowledge validation and refinement.

These arguments seem valid for the actual project as it should provide some insights in the relationships within the data and not only lead to high accuracy. Therefore I will rely on Logistic Regression as well as on Decision Trees and Random Forests because their models can be interpreted and provide in addition valuable insights in the impact relations between the variables.

3.1 Representativeness of the Collected Data

As we learned from chapter 2 the students had the possibility to decide if their actions in ILIAS should be tracked anonymously or with pseudonyms. A decision for pseudonyms would imply that the events in the data set (produced by the actions taken in ILIAS) can be connected to the results of the homework assignments and the final exam. But because of the low rate of pseudonymously tracked students, we need to make sure that this group of students is not a self-selected, statistically different subset from the population. The pseudonymous students should have the same characteristics as the anonymous students.

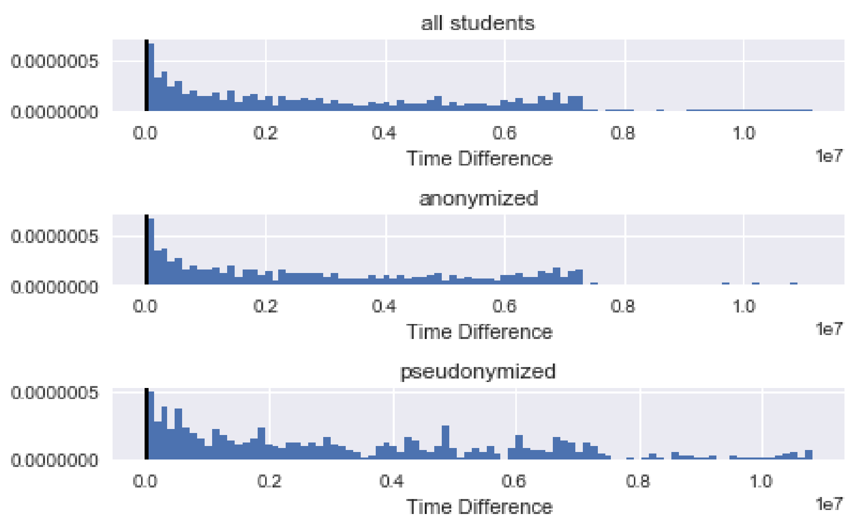

Figure 3.1: Lecture: Time Difference between Upload of File and Student Interaction

The figure 3.1 shows the frequencies of time differences between uploading and accessing a file in the lecture folder of "Praktische Informatik I". From top to bottom we see the frequencies for all students, anonymized and pseudonymized students, the unit of the x-axis is "second".[1] The histogram for the pseudonymized students looks less constant, but we have to keep in mind that this group includes only 26 students, with 215 students remaining anonymous. With the mass smoothing out single extreme values, it is only natural that the smaller group shows more peaks in the histogram.

We find this pattern repeated for the other variables under investigation (access

[1]This unit was selected because of plotting feasibility.

relative to the exam date, weekday, hour of day). It would be out of scope to show all of these histograms, but their patterns accord with the results previously found in the literature (Sheard, 2011a, p. 317) and some additional figures can be found in the Appendix. Additionally, table 5.1 in the Appendix shows the means of these variables, separated by groups. For each variable I applied the so-called Mann-Whitney-U test,[2] which is also used in previous studies for similar data (Sheard, 2011a). Again, detailed results can be found in table 5.1, the overall picture looks ambiguous. For three of the four analyzed variables (as well for the tutorial as for the lecture data), the differences between the groups seem significant, whereas for the exam results we do not find any significance. The absence of a clear picture may come from the dramatic discrepancy in the size of the samples (as noted above).

3.2 Team Composition and Homogeneity

Having the data about working behaviors of students at hand, a short additional analysis of the tutorial teams could be interesting for lecturers as well as for the students themselves. Some would expect to find the teams formed by equally good students, or weaker students to seek the company of better students to make passing the course more easily. In the following paragraphs, we will see if any of the named expectations is found in reality.

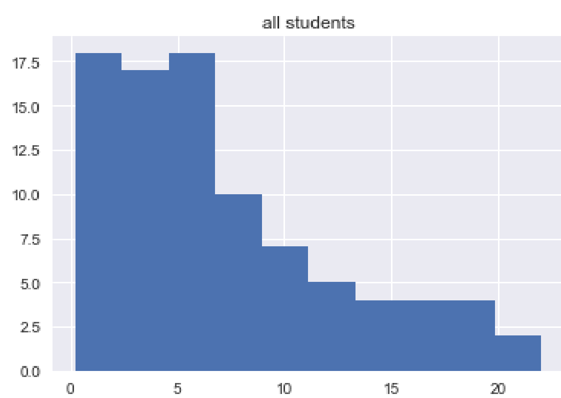

Figure 3.2: Absolute Grade Difference to Team Average (Without Alone Working Students

[2]The Mann-Whitney-U test is (in contrast to the "classical" t-test) a non-parametric test, i.e. it does not assume any shape in the data. In its essence: It compares two samples and analyzes if these stem from a population with the same distribution. (Sheard, 2011b, p. 39)

3.2.1 Is there a tendency for better students to form groups together?

The histogram 3.2 shows the distribution of differences between the team members and the team average. In this case, the students who were working alone were excluded from the graph. As a first impression we see an almost curved shape with a lot of teams very close and some teams wider apart. This distribution shape supports the first formulated expectation more because we find a lot of students up to 5 points from the team average (that means 10 points to their team member).

3.2.2 Are the group members very alike?

Figure 3.3 shows the students' grades compared to their team members. As the allocation to member groups 0 and 1 was done randomly, all points are doubled (light blue, darker blue) to avoid an accidental trend in the distribution of points. When we look at the plot, we see no clear trend in the data points. There seems to be a tendency to the top right, so high scoring students seem to work more often with equally high scoring students, while the positioning of points to the left bottom becomes less dense. However, the correlation coefficient between the team members is about 0.28, which is considered to be a low correlation value (ranges from 0 to 1) and therefore does not show any truly significant connection within the teams. Because of the randomness of member group allocation a t-test or any other statistical test does not make sense.

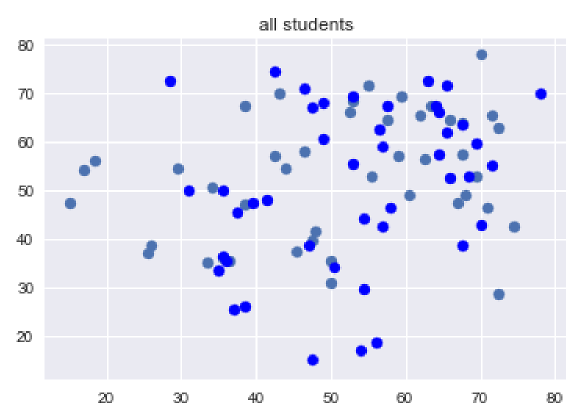

Figure 3.3: Student Grades Plotted against their Team Member

3.3 Correlation and Pattern Detection

So far we have seen that there is no serious pattern in students forming teams, and we have seen that we can't be sure about our self-selected pseudonymous students being statistically not different from the population. Now we want to have a look at the underlying question, where we started this paper originally.

3.3.1 Can we predict the success of students attending the course and how early can we do it?

Because only a few students decided to give their data with pseudonyms, I did the first analysis only with exam and homework data. The goal was to predict the success based on the weekly homework assignments. The predictive analyses were done in varying ways.

First, I started with all students available and reduced the number of independent variables step by step. The first step included homework number 1 to 5, semester, tutor, subject, and degree. The second step used faculty instead of subject because I thought that subject may be too fine granular to have an impact. In the third step, I excluded tutor, subject as well as faculty from the data. Each time I included homework number 1 to 5 because the middle of the semester would be a good point to analyze the progress of the students, while it would not be too late to change something in the learning and working behavior. These three data sets were randomly split into training and testing, two independent splits created to prevent overfitting. As a baseline, I chose the Naive Bayes algorithm, its outcomes will be compared to Logistic Regression, Decision Trees, and Random Forests. I selected these algorithms as their models provide some insights into the weighting and importances of the variables included. As we want to understand the causes and relationships between learning, working, and grades, these insights are crucial to our analysis. For Decision Trees and Random Forests, I did some parameter tuning. With a grid search, I checked for the best value of depth (DT) and the best combination of depth and the number of estimators (RF).

	Data Step 1	**Data Step 2**	**Data Step 3**
Naive Bayes	0.58	0.63	0.61
Logistic Regression	0.63	0.61	0.66
Decision Tree	0.68	0.71	0.71
Random Forest	0.76	0.79	0.79

Table 3.1: Accuracy-Results of First Approach

Table 3.1 shows the overall results of all classifiers and data sets. This table reports accuracy values, but the F1 values are quite comparable, this can be seen in table 5.2 in the Appendix. In general, the Random Forest classifier was able to produce the best prediction results. Even its weakest result of 76 % accuracy could not be reached by the best result of any other classifier. The best value of nearly 80% accuracy was achieved with data sets 2 and 3, i.e. with the smaller data sets. For all classifiers, these data sets seem to work better. This is indeed very interesting because data set 3 includes fewer variables (less information) but works equally well than data set 2 and better than data set 1. We will see how these results perform in comparison to the other analysis steps.

Please always keep in mind that the size of the training set is 212 students and the testing set therefor only contains 38 records. One changed predicted student leads to an accuracy change of 2,5%.

Because it would be necessary for a lecturer to be informed about the bad performing students a recall as high as possible (of these "failing" students) would be great. I weighted these failing students within a Random Forest classifier by oversampling them (in different rates) and compared the recall values and the precision rates. The results can be seen in figure 3.4. As we know there is a trade-off between high recall and high precision, but the best oversampling rate value seems to be around 5 and 10.

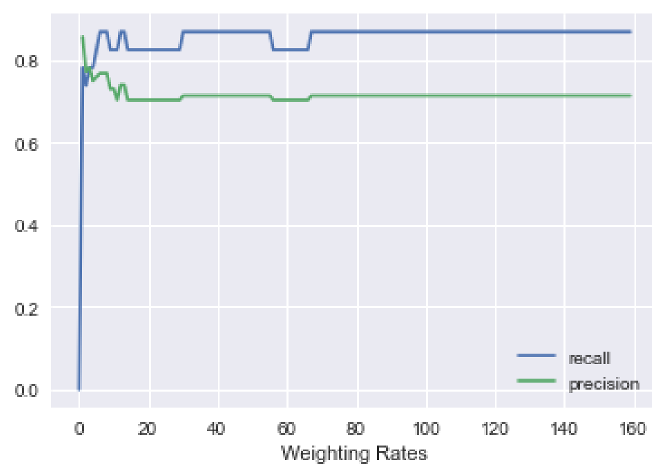

Figure 3.4: Recall and Precision Compared at different Oversampling Rates

The **second** approach was to investigate rates at all weeks during the semester. How precise are our predictions when we have homework information from the first week? At which point do we have "enough" information to predict best? For this task, I used the "smallest" data set at hand because we found out in the previous task that it was the best performing (compared to the number of variables used). This data set was copied ten times, from each copy I deleted more and more weeks beginning at the end of the semester until I had only the first week left. These eleven sets were split in training and testing and fed into a Random Forest classifier, which performed best in the previous task.

Figure 3.5 shows that we are able to predict the success with more than 80% accuracy directly after semester week 2. This value is not reached with semester weeks 3-8 added, only if we include week 10 and 11, the last two homeworks in the semester, the results exceed week 2. This is not quite suprising, because at this point in time it should be clear which students are admitted for the exam and which not (they fail automatically!). But, and this is the most interesting finding, we classify 4 out of 5 students correct in passing and failing after the second round of homeworks, which could be a valuable information for the lecturer at the beginning of the semester to get a clue which students maybe need additional help.

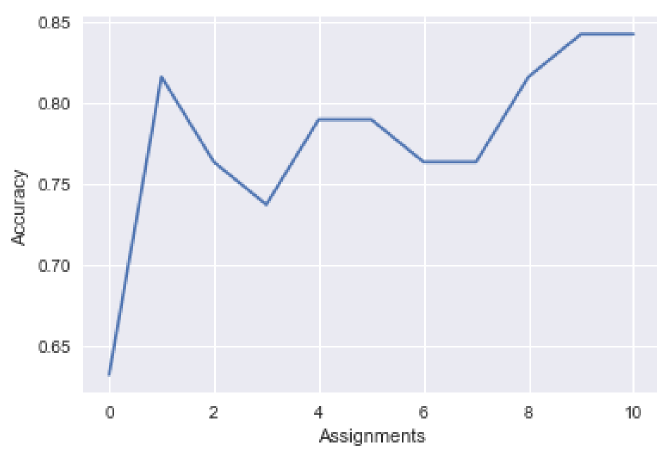

Figure 3.5: Accuracy Development over different Assignment Weeks

It is not hard to predict the students correctly who dropped out because they did not reach the minimum of points during the semester, and weren't admitted to the exam. In the **third** step I want to predict the results only of those students who attended the exam and tried to write it (they got at least some points). This may also take some difficulties away because some of the students have 0 points in the first exam to write it on the second take, which took place several weeks after the first try. They are marked as "failed" but may have done this on purpose and were good students, nevertheless. Unfortunately, I have no access to the results from the second take.

	Data Step 1	**Data Step 2**	**Data Step 3**
Naive Bayes	0.58 (+0.00)	0.77 (+0.14)	0.85 (+0.24)
Logistic Regression	0.73 (+0.10)	0.73 (+0.13)	0.77 (+0.11)
Decision Tree	0.77 (+0.09)	0.77 (+0.06)	0.77 (+0.06)
Random Forest	0.81 (+0.04)	0.88 (+0.10)	0.85 (+0.06)

Table 3.2: Accuracy Results of Third Approach (Comparison to First Approach in Parentheses)

As expected, the results get better. For some classifier/data set combinations we gain a plus of more than 20% accuracy. This increase makes the Naive Bayes classifier for the smallest data set equally good as the Random Forest classifier, but it also pushes the Random Forest classifier to 88% accuracy for data step 2. That means this model classifies nearly 9 out of 10 students correctly into passing and failing if they at least try to write the exam.

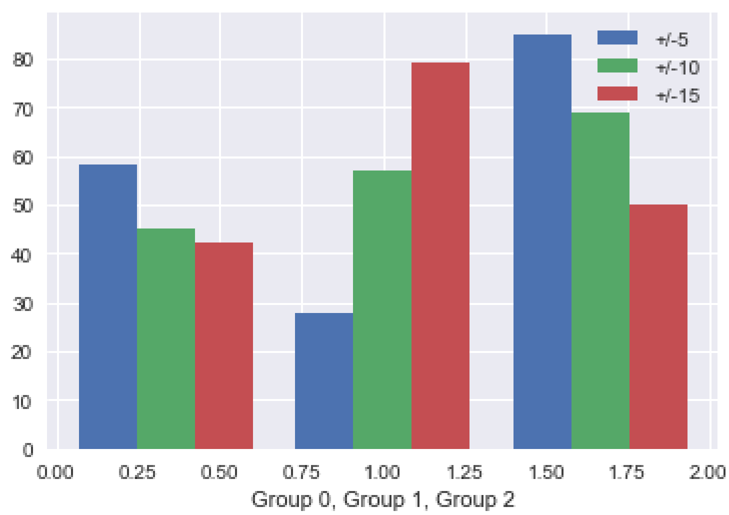

Figure 3.6: Students Distributed in Groups: 0, 1, 2

Finally I did a last approach not to categorize the students into failing and passing, but in three classes: clearly failing, critical, and clearly passing. As we all know, between failing and passing an exam is a clearly drawn line and a few points can all make the difference in the end. For a lecturer, it would be much more appropriate to know which students need help most urgently. Krumm and colleagues also use this categorization in their "Student Explorer", they call the three classes

"Engage, Explore, Encourage"[3] (Krumm et al., 2014, p. 109). When we create this three classes we use three different approaches to pre-classify the students for the target variable. The critical group (between clearly passing and clearly failing) spread either +/- 5, 10 or 15 points from the passing threshold of 42.5 points. The distribution of students within these groups can be seen in figure 3.6. I created different spreads because the group sizes differ dramatically and depending on the group sizes, the algorithms could perform significantly worse or better. This analysis was done again with the "smallest" data set again, only containing the homework assignments as well as semester and the anticipated degree.

	42.5 points +/-	**Group 0**	**Group 1**	**Group 2**	**Average**
Naive Bayes	5	0.95	0	0.8	0.67
	10	0.95	0.5	0.88	0.82
	15	0.95	0.4	0.78	0.72
Decision Tree	5	0.82	0.2	0.8	0.67
	10	0.8	0.5	0.88	0.76
	15	0.8	0.58	0.46	0.62
Random Forest	5	0.95	0	0.8	0.67
	10	0.75	0.53	0.86	0.75
	15	0.8	0.48	0.5	0.6

Table 3.3: F1-Results of Different Classifier for 3-Level-Prediction

Table 3.3 shows that the highest overall F1-score was produced by the Naive Bayes classifier, working with a critical group +/- 10 around the threshold. The best result for the critical group 1 was achieved by the Decision Tree classifier with +/- 15. The major problem for this analysis was the small number of records within the +/- 5 range. The classifiers could not predict more than 0.2 F1-value with this target variable. With the +/- 10 target, only the Naive Bayes classifier could work best, while both Decision Tree and Random Forest need the bigger size of +/- 15 critical target group (while getting worse for group 2). This may explain the best overall score of Naive Bayes.

[3] "[W]ether instructors should encourage students, explore their progress in more detail, or engage with students to asses possible academic difficulties." (Pardo, 2014, p. 32)

3.3.2 Does the usage pattern of ILIAS have an influence on the students' performances?

Unfortunately, because of the low pseudonymous student rate, we are left with 17 (lecture data) and 16 (tutorial data) student records which had both, ILIAS data and exam grade, available. Within these small data sets, only 2 students each did not pass the exam, which means that this group is very underrepresented. I tried to multiply these records to make the classes more balanced, but this lead to overfitting and 100% accuracy for nearly all classifiers. So I ended up with splitting the data into training and test sets, where the test sets just contained 5 and 6 records, but one of them was definitely a failing student.

	Lecture		**Tutorial**	
	Accuracy	F1	Accuracy	F1
Naive Bayes	0.83	0.76	0.8	0.71
Logistic Regression	0.66	0.67	0.8	0.71
Decision Tree	0.83	0.76	0.8	0.71

Table 3.4: Accuracy and F1-Results for Models with ILIAS Data

In table 3.4 we see: All classifiers except for Logistic Regression failed to capture the failing student (that's why the results are all equal). Only Logistic Regression predicted one truly passing student as failing. When we have a closer look at the models, for example, the model from the Decision Tree classifier, we can see that it uses only one (continuous) variable to separate its decisions. This variable changes randomly when the model is re-computed again and again. It fails to capture patterns in the data, maybe because there are none (because we gathered wrong variables), maybe because it has too few instances to build a robust model.

3.3.3 Which variables have the highest impact and how would a change in these variables affect the grades in the end?

We can look also at the variable importances of the previously calculated models. Figure 3.7 pictures the importances from the Decision Tree and the Random Forest classifier. Both decide their predictions based upon single variables, these decisions can be combined to importances, which sum always in total up to 1. We have to keep this in mind because data sets 2 and 3 contain fewer variables to distribute the importance factors.

Figure 3.7 shows that assignments 1-5, as well as the semester, are highly used as variables for decision making. Especially assignment 2 seems to be critical (this supports what we have seen in figure 3.5). It is followed by assignment 5 and the semester. But these first variables were used in every data set. From the other variables, some tutors, as well as the degree difference, seems to be important. But, and this might be the explanation why data set number three (dispite being the smallest) gains the best results with the classifiers, the importance values are only marginal when compared to the values of the assignments and the semester variable.

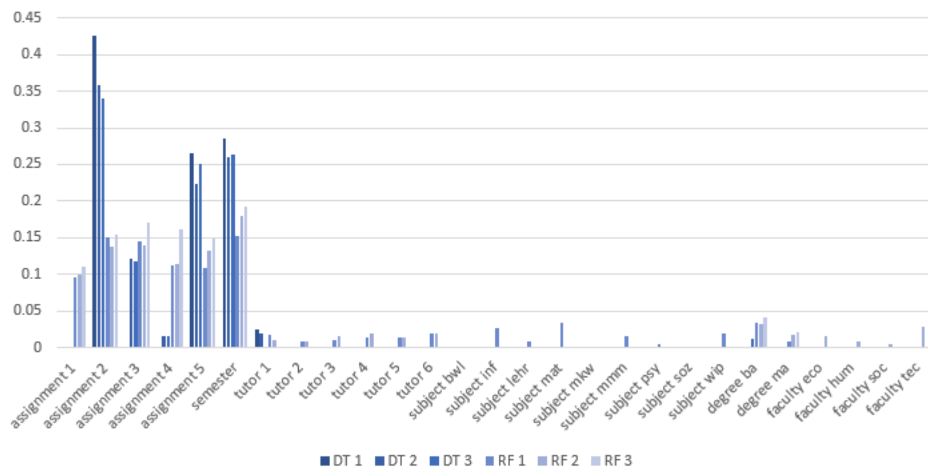

Figure 3.7: Variable Importances of Decision Tree (DT) and Random Forest (RF)

The Logistic Regression coefficients in figure 3.8 show a contrasting picture. Their absolute values are not easy to interpret [4], but we can say something about their direction and their relative size. We only have to keep in mind, that the categorical variables, such as tutor, subject, degree and faculty, have each a reference category, which was left out for the regression calculation. All coefficients have to be interpreted in relation to this reference category, for example: An undergraduate (bachelor) student is statistically more connected to pass the course than a postgraduate (master) student. The reference categories are the following: tutor 1 for tutors, subject "mkw" (media and communication science) for subjects, master for degree and romanistic faculty for faculty.

[4]An one unit increase in the dependent variable is connected with the logarithmic odds ratio of the respective independent variable.

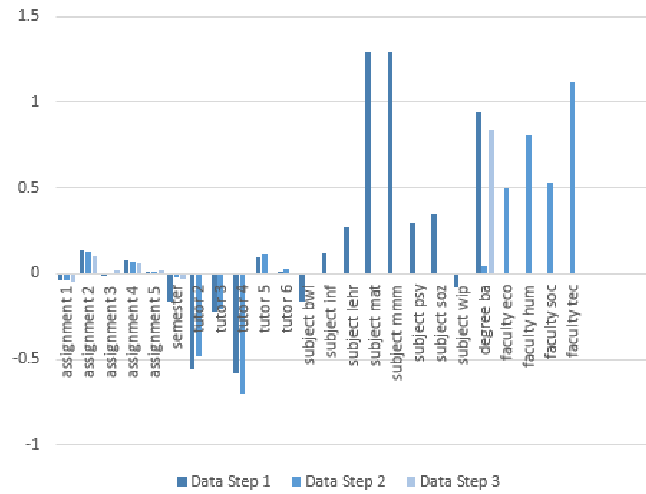

Figure 3.8: Regression Coefficients of Logistic Regression

Only six variables in total show a negative relationship with the target variable "pass" (are therefore connected rather with failing than with passing): tutors number 2 to 4, as well as the semester and the subjects "bwl" (business) and "wip" (business education). The highest positive coefficients are subjects "mat" (mathematics) and "mmm" (management)[5], the degree and the all the remaining faculties (if uncluded in the model)[6]. The assignment variables don't seem to be highly connected to the target variable at all, and, in contrast to the importances from Decision Tree and Random Forest we have seen above,the semester has a rather negative relationship to passing the exam than a positive one.

[5]Always in comparison to subject "mkw".
[6]In comparison to the romanistic faculty.

Chapter 4

Discussion

In the past chapters, we have analyzed the difference between anonymous and pseudonymous students, the patterns of student groups, as well as possible impacts on the performance of students in the exam. The analysis showed that semester, degree, but mostly the weekly assignments are good indicators for passing and failing. Maybe someone could build upon this and introduce for example weekly questionnaires for students, about how he or she liked the course so far, where he or she struggles, and how much/what he or she learned. Experiments, with for example "nudging" (Pardo, 2014, p. 32), could lead to further insights[1]. Also, as (Costa et al., 2017) have shown, there is a high potential in additional preprocessing (at least for distance education data), which could be done also for this project.

We have seen that one course with the default setting "anonymous" and the option to activate the account to "pseudonymous" is definitively not enough to apply data mining techniques on interaction data. As far as I know it is not possible to grand those "activated" accounts benefits to raise the overall percentage, because the agreement to give away personal data has to be completely free and voluntary. But maybe it could be possible to carry on this analysis with data from following semesters (even though this can be hard because of changing lecturers, tutors, circumstances in general).

It would also be very interesting to know more about the contents and topics first of the links and files used, and second of the written assignments. Some could draw links between files - assignments - single questions in the exam, and maybe what file somebody missed to answer question X correctly. We could measure knowledge in terms of "topics" and weight them for the exam differently.

Besides content analysis previous works show far more possibilities which can

[1]This approach was considered by the responsible Chair of Economic and Business Education - Learning, Design and Technology, the data gathering has already started.

be applied here in future work: We could automatically cluster students based on their online behavior (Lee, 2018), analyze session patterns from the interaction data (Sheard, 2011a), create an "action library" with "sets of sequential event traces" (Zhou et al., 2011, p. 113) that lead to higher level information, mine sequential patterns directly (Tang et al., 2018), or we could predict "deserters" (those students, who start a course but drop it after a specific amount of time or do not start to work for it at all) to encourage them to work (Aguilar et al., 2018).

This topic leaves so many options to investigate further in research, but we have to keep in mind: The interactions analyzed in the previous chapters only represent a minor fraction of real course interaction (Pardo, 2014, p. 23). We have no idea about what happened in the classroom, tutorials, via email or in person. ILIAS support was only additional help for the students (as the lecture was considered as a "blended" course), so a student could have passed with 1.0 without looking in ILIAS at all. Therefore, when we draw conclusions from our analysis, we have to remember that our interaction data are only tiny bits of actions in total and that our results are not determined relationships but tendencies.

One last remark: Do not forget that passing, failing, grading is not always a valid indicator for how much a student learned or how much effort he or she spent in working and learning, neither it is a proxy for fun or enjoying the course!

Chapter 5

Appendix

5.1 Additional Tables and Figures

Tutorial	Anonymous	Pseudonymous	Difference Significant
Time Until Download	33 days, 03:11:26.63	31 days, 20:17:48.44	no
Access Relative to Exam Date	42 days, 14:44:12.91	45 days, 10:45:41.81	yes
Weekday	3.65	3.46	yes
Hours/Day	14.57	14.61	yes

Lecture	Anonymous	Pseudonymous	Difference Significant
Time Until Download	36 days, 06:42:16.64	37 days, 15:53:54.20	yes
Access Relative to Exam Date	37 days, 04:40:44.96	36 days, 06:45:03.80	no
Weekday	3.79	3.5	yes
Hours/Day	14.85	15.03	yes

Exam Results	Anonymous	Pseudonymous	Difference Significant
Points	33.98	38.63	no
Grade	2.18	3.04	no

Table 5.1: Comparison: Means of Anonymous and Pseudonymous Students

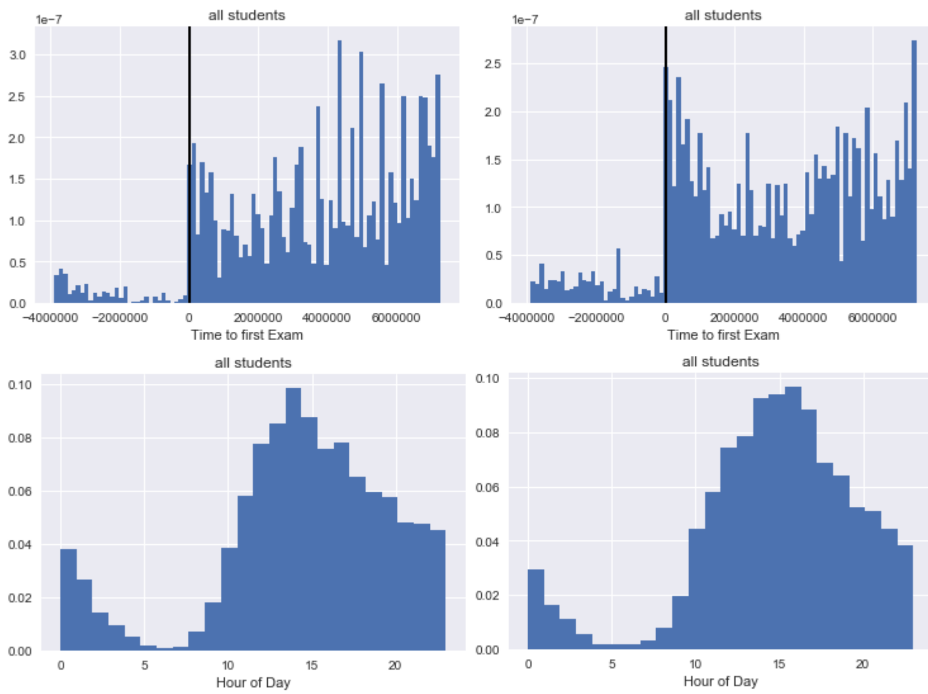

Figure 5.1: Comparison of Tutorial (left) and Lecture (Right) Usage Rates

	Data Step 1	**Data Step 2**	**Data Step 3**
Naive Bayes	0.58	0.63	0.6
Logistic Regression	0.63	0.61	0.64
Decision Tree	0.61	0.71	0.71
Random Forest	0.77	0.79	0.79

Table 5.2: F1-Results of First Approach

	Data Step 1	**Data Step 2**	**Data Step 3**
Naive Bayes	0.51	0.76	0.84
Logistic Regression	0.72	0.72	0.75
Decision Tree	0.75	0.75	0.75
Random Forest	0.8	0.88	0.85

Table 5.3: F1-Results of Thrid Approach

5.2 References

[Aguilar et al., 2018] J. Aguilar et al. 2018. Specification of the autonomic cycles of learning analytic tasks for a smart classroom. *Journal of Educational Computing Research*, 0(0):437–469.

[Asif et al., 2017] R. Asif et al. 2017. Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, (113):177–194.

[Baker and Inventado, 2014] R. S. Baker and P. S. Inventado. 2014. Educational data mining and learning analytics. In J. A. Larusson and B. White, editors, *Learning Analytics: From Research to Practice*, chapter 4, pages 61–78. Springer, New York.

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

[Costa et al., 2017] E. B. Costa et al. 2017. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73:247–256.

[Hämäläinen and Vinni, 2011] W. Hämäläinen and M. Vinni. 2011. Classifiers for educational data mining. In R. Cristóbal et al., editor, *Handbook of Educational Data Mining*, chapter 5, pages 57–74. CRC Press, Boca Raton.

[Krumm et al., 2014] A. E. Krumm et al. 2014. A learning management system-based early warning system for academic advising in undergraduate engineering. In J. A. Larusson and B. White, editors, *Learning Analytics: From Research to Practice*, chapter 6, pages 103–122. Springer, New York.

[Lee, 2018] Y. Lee. 2018. Using self-organizing map and clustering to investigate problem-solving patterns in the massive open online course: An exploratory study. *Journal of Educational Computing Research*, 0(0):1–20.

[Livieris et al., 2018] I. E. Livieris et al. 2018. Predicting secondary school students' performance utilizing a semi-supervised learning approach. *Journal of Educational Computing Research*, 0(0):1–23.

[Masip et al., 2011] D. Masip et al. 2011. Capturing and analyzing student behavior in a virtual learning environment: A case study on usage of library resources. In R. Cristóbal et al., editor, *Handbook of Educational Data Mining*, chapter 24, pages 339–351. CRC Press, Boca Raton.

[Pardo, 2014] A. Pardo. 2014. Designing learning analytics experiences. In J. A. Larusson and B. White, editors, *Learning Analytics: From Research to Practice*, chapter 2, pages 15–38. Springer, New York.

[Scholes, 2016] V. Scholes. 2016. The ethics of using learning analytics to categorize students on risk. *Education Tech Reserach Dev*, (64):939–955.

[Sheard, 2011a] J. Sheard. 2011a. Analysis of log data from a web-based learning environment: A case study. In R. Cristóbal et al., editor, *Handbook of Educational Data Mining*, chapter 22, pages 311–363. CRC Press, Boca Raton.

[Sheard, 2011b] J. Sheard. 2011b. Basics of statistical analysis of interactions data from web-based learning environments. In R. Cristóbal et al., editor, *Handbook of Educational Data Mining*, chapter 3, pages 27–42. CRC Press, Boca Raton.

[Tang et al., 2018] H. Tang et al. 2018. Time really mattes: Understanding the temporal dimension of online learning using educational data mining. *Journal of Educational Computing Research*, 0(0):1–22.

[Wanli et al., 2015] X. Wanli et al. 2015. Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, (47):168–181.

[Zhou et al., 2011] M. Zhou et al. 2011. Sequential pattern analysis of learning logs: Methodology and application. In R. Cristóbal et al., editor, *Handbook of Educational Data Mining*, chapter 8, pages 107–121. CRC Press, Boca Raton.

5.3 Repository

Extracts from the data sets, as well as the code that produced all calculations, tables and figures, can be found in my Git repository: Individual Project - Mining ILIAS Data (https://github.com/LarissaHa/individual-project/)

5.4 Statement of Authorship

I hereby declare that the paper presented is my own work and that I have not called upon the help of a third party. In addition, I affirm that neither I nor anybody else has submitted this paper or parts of it to obtain credits elsewhere before. I have clearly marked and acknowledged all quotations or references that have been taken from the works of others. All secondary literature and other sources are marked and listed in the bibliography. The same applies to all charts, diagrams and illustrations as well as to all Internet resources. Moreover, I consent to my paper being electronically stored and sent anonymously in order to be checked for plagiarism. I am aware that the paper cannot be evaluated and may be graded "failed" ("nicht ausreichend") if the declaration is not made.

LHaas
Larissa Haas
September 7, 2018