Predicting Dropout in an Online Degree across Two Institutions - A Case Study

Agathe Merceron[1[0000−0003−1015−5359]], Teodora Dogaru[1], Raj Waghela[1], Tabea Sophie Harper[1], and Petra Sauer[1]

Berliner Hochschule für Technik, Luxemburgerstrasse 10, 13353 Berlin, Germany
{merceron,teodora.dogaru,sauer}@bht-berlin.de

Abstract. In this paper, we investigate predicting students at risk of dropping out using academic data from an online study program offered at two different universities. In this program, most students work full-time and thus study part-time. We use the algorithms decision tree and logistic regression, which have given good results in other works and are interpretable. The decision trees or the coefficients produced by logistic regression can be shared with stakeholders to identify courses that impact dropout the most and that could be considered for additional support. Our results are comparable to the results that others have obtained when considering face-to-face study programs. We propose a preliminary approach to identify courses that impact the prediction the most. Our results also show that merging the data from the two universities is not helpful for the prediction. This might be due to differences in the regulations that impact how students study.

Keywords: Predicting Dropout · Online Study Program · Explainable Models · Decision Tree · Logistic Regression.

1 Introduction

Due to the high rates of students dropping out of their studies in higher education, predicting whether students will drop out has been intensively researched, see [2,4,5,17,27] for a few examples. In Germany for instance, 28% of the students dropped out of their bachelor study program in 2020 [13]. When predicting whether students will drop out of a degree program, till now, the research has focused on classical so-called face-to-face study programs, as do the works mentioned above. This could be because these classical degree programs are the most common. In Germany for example, only 4.9% of the degree programs in higher education in 2024 are distance learning programs, including the online degree program considered in this work [14]. We investigate predicting dropout in an online degree program called "Computer Science and Media" (CM) in which most of the students work full-time and, thus, study part-time. This program was introduced in 2001. Our work can be seen as a replication study using known algorithms to predict dropout with different data, data from an online study program instead of a face-to-face one as commonly addressed in the literature.

Note that we predict whether students will drop out of an online study program and not of a (massive open) online course.

This six-semester undergraduate online degree program is offered in a network of German public universities of applied sciences. Casual surveys show that many of the students enrolled in this online degree work full-time or have small children or relatives they take care of (note that universities do not store data on this aspect). Being online, this degree gives students great flexibility as they can study from their place at their own pace. Most students enrolled in this degree would not be able to attend a classical degree program with face-to-face teaching and a fixed schedule. As we will see later in Section 3, the number of students dropping out of their studies is higher than in classical face-to-face study programs, where dropping out means abandoning the program before getting the degree. In our context, dropping out does not necessarily mean failure, see [28]. For example, some students can get a better job thanks to the skills and knowledge they have acquired after studying a few courses and, thus, quit the program before obtaining the degree.

When enrolling in this program, students choose one university of the network. The syllabus, the credits, and the standard course material available in the learning management system (LMS) are the same for all universities offering this online degree. The standard course material contains reading material, videos, animations, quizzes, worked-out exercises as well as tests; the number of these elements may vary, depending on the courses. Each university appoints its teaching staff; except for a few optional courses, the teaching staff is not shared among the universities. Like in face-to-face teaching, teachers can conduct their online courses in their own way. For example, they can encourage the usage of the forum, or not, or, in addition to - or even instead of - the standard material of the course provided by the network, they are free to use any teaching material they like as long as they follow the syllabus. Teachers offer weekly or bi-monthly synchronous online meetings, as well as two to four in-class meetings in a semester. Depending on the courses, these meetings can be mandatory or optional. Regulations may vary among the universities. For example, some universities require the students to register for the exam in a course shortly before the end of the semester, while others do not: any student enrolled in the course is allowed to take the exam. As another example, some universities have an upper limit on the number of times students may enroll in the same course, while others do not.

The degree program considered in this work has six semesters if students study full-time. However, students are free to study at their own pace. Because of other obligations, students usually study part-time as will be substantiated in Section 3. This means that many of them do not enroll in six courses each semester - as scheduled in the study handbook - but less. Further, the program has a limited number of seats, resulting in a smaller number of students than considered in other works, see [2,4,5,6,17,27], which also predict students' dropout. Compared to those works, our dataset is small. Thus, this research is also a contribution to which extent machine learning approaches help predict dropout in the case of small data sets.

We are interested in predicting students' dropout using interpretable machine learning algorithms, interpretable in the sense of Miller p.14 "the degree to which an observer can understand the cause of a decision" [18]. Indeed, this research is part of a bigger project involving different stakeholders who must understand the results. Therefore, we use primarily the algorithms decision tree and logistic regression, see [19], to make the prediction. These algorithms have given good results in other works [9,27].

In our work, we do not build a new dataset and do not collect data directly from the students. Instead, we use the data that any university in Germany must store about its students according to German law. We thus reproduce the data acquisition approach of other works, like [2,4,5,6,17,27] to predict dropout. Using this readily available data makes our work much more easily repeatable over the years and more easily reproducible for others. In Germany, compared to universities in the USA, universities store limited demographic data. For example, according to the law, German universities must not store information related to ethnicity, the obtention of a grant, the economic situation of the parents or the students, or whether students work besides their studies. Likewise, we have no information about the motivation of the students, which is an important factor influencing dropout, see [24,12].

Even when utilizing only the data that universities must store about their students, different kinds of student data can be used to predict dropout. The work of Berens et al. done in a German context shows that academic data are far more important than demographic data in predicting dropout [4]. Therefore, in this research, we explore predicting dropout using only the academic data stored by the university which essentially is the courses students enroll in each semester and the marks they get in the exams. The reasons for this choice are the work of Berens et al. [4] mentioned above, the good results obtained by others like Manrique et al. [17] who use also only academic data, and a responsible handling of data protection for students: less data make inferences about individual students less likely.

As argued by Manrique et al., different kinds of features can be calculated from the academic data [17]. In this work, we choose to use what Manrique et al. call "a local feature set", which means that we use only the courses students enroll in and their grades. We do not calculate global features such as the number of failed courses, the average mark, and so on. The reason we do so is that we would like to see whether specific courses are important to predict dropout. This knowledge could be useful to develop guidelines or interventions that might prevent dropout. This information will be conveyed to several stakeholders, like the heads of the program to possibly reflect on the structure of the syllabus, to the teachers, and the pedagogical team for considering additional help, like dashboards for teachers to follow their students better or dashboards for students to support them in their learning and self-regulation process.

We have obtained the academic data from two universities of the network. This research complies with the data protection and security regulations of our university and with GDPR (General Data Protection Regulation of the European Union).

Our research questions are:

RQ1: What is the performance of the classifiers ‘decision tree’ and ‘logistic regression’ when predicting whether students are at risk of dropping out of the online degree? Which courses are important for the prediction?

RQ2: Does predicting dropout for each university separately give a better performance than merging the two academic datasets?

2 Related Work

Researchers have used diverse data sources to solve the task of predicting dropouts: pre-entry data like the grade of entrance degree, demographic data like gender or age, and academic performance data like course grades and course enrollments. Manrique et al. [17] have achieved good scores using academic performance data only. Aulck et al. [2], Berens et al. [4], and Cai and Fleichhacker [5] have shown that adding demographic data to the academic data hardly improved the results. Such results were confirmed in the study of Cohausz et al. [8]. Further, Baker et al. [3] and Cohausz et al. [8] argued against using demographic variables as predictors. Taking these works into account, and because of data protection, we have considered only data related to the academic performance of students to build predictors in this work.

Various students’ representations can be calculated from academic performance data to predict whether students will drop out. In the work of Manrique et al. [17], three student representations were distinguished, which the authors referred to as global features, local features, and time series. The local feature set contains only courses and their grades, which are directly part of the academic data. The global feature set contains features calculated from the academic data such as average grade or the number of failed courses; the generation of this set requires a feature engineering step. Dekker et al. [9] and Kemper et al. [16] have used a mix of local and global academic performance features while the research by Aulck et al. [2] and Berens et al. [4] has been limited to global features. An advantage of models based on global features is that they can be trained for multiple study programs together; however, Manrique et al. have reported that all their models trained with the local feature sets had a better performance than the other models [17]. In contrast, Wagner et al. [27] did not obtain better results with the models trained with a local feature set but the models trained with the global feature sets did not predict at-risk students across three degree programs equally well. Manrique et al. stress that using time series results in more complex models [17]. In this work, we considered only the approach with local feature sets as we sought to identify which courses impact the prediction the most.

As already mentioned, we used the algorithms decision tree and logistic regression because they are explainable and the visual representation of a tree or of a list of coefficients is understood by stakeholders like heads of study programs. These algorithms have been used and have given good results in other works [9,16,27]. Related work has taken different approaches to improve their models by preprocessing the training data. Kemper et al., [16], for example, have removed unpopular exams with fewer than 15 scores each for dropouts and graduates. Manrique et al. in contrast, have included only mandatory courses based on the assumption of at least 20 enrollments per semester [17]. In both cases, the same approach has been used for all programs and semesters. Wagner et al. developed a program-specific approach to find the courses that might be critical for predicting dropout [27]. Because we predict dropout at the end of the 1st, 2nd, and 3rd semesters, we have used all courses of the 1st, 2nd, and 3rd semesters, which are all mandatory courses. As already mentioned, because most of our students study part-time, they do not enroll in all the courses planned in a semester, resulting in many missing values. Our dataset is imbalanced, as more students are dropping out of the study program than graduating. Kemper et al. and Manrique et al. [16,17] have used SMOTE to balance the ratio using synthetically generated data sets because dropouts and graduates were not equally distributed in the data. Aulck et al. in contrast, have intentionally trained the models using data "in its original, unaltered form" [2], p.5. We follow the approach of Aulck et al. [2] and Cai and Fleichhacker [5] and use the original data as it is, as we want the decision trees and the coefficients to reflect reality.

Regarding the evaluation of the models, we have followed the approach of [27] and considered balanced accuracy as our key metric. Balanced accuracy is the mean of the true positive rate and the true negative rate and gives equal attention to the correct prediction of dropouts and graduates. The main reason is that our data is very imbalanced.

One aim of predicting whether students are at risk of dropping out is to develop some kind of early warning system like in [4,6,17,21]. Schneider et al. report on contacting per email students who were predicted at risk of dropping out with a probability of more than 50% and inviting them for personal counseling [21]. This had almost no consequence because students did not follow the invitation, except for a handful of them. They could not detect any effect on the subsequent behaviors of students except in engineering study programs: students having received such an email tended to drop out earlier and more. A similar observation has been reported by Jayaprakash et al. [15] in courses where the early warning system signals [1] had been implemented: on one hand, the performance of students who completed the course improved, on the other hand, students dropped out of the course more, which had not been intended. It is not clear whether students wish to receive any kind of early warning as reported in [25] or [20]. Wagner et al. use dropout prediction to evaluate a course recommender system intended to support students at risk of dropping out while enrolling for courses at the beginning of the semester [26]. In this research, we aim to show different stakeholders which courses impact the decision of the models the most so that

Table 1. Number of students at the end of their 1st (S1) 2nd (S2) or 3rd (S3) semester of studies who finally graduated (G) or dropped out (D) in the universities B and T.

	B			T		
	S1	S2	S3	S1	S2	S3
G	28	27	26	54	54	54
D	233	156	114	423	256	168
Sum	261	183	140	477	310	222

they might develop interventions, for example, changes in the curriculum, in the course material, in the onboarding event at the beginning of the 1st semester, or the use of a dashboard that could reduce the number of students dropping out.

We contribute to the field of responsible knowledge discovery in education:

– by considering an online study program offered in two universities in which most of the students study part-time. Most of the related works consider classical face-to-face study programs.

– by comparing the performance of the models built with data from each university with the performance of the model built with a merged dataset including data from both universities. Most of the related works like the ones cited in this paper consider study programs from a single institution.

– by proposing a novel though still preliminary approach to summarize the explanations of the different models transparently and understandably for stakeholders.

3 Data and Methodology

We introduce the data used in this study and we describe the methodology used to predict dropout and evaluate the performance of the prediction.

3.1 Data

We collected and anonymized data from two universities of the network, B and T. Even though the six-semester undergraduate program "Computer Science and Media" was established in 2001, we ignored in this study data stored before the winter semester of 2014 because of major course changes in the curriculum. We obtained from T data till the winter semester of 2022 and from B data till the summer semester of 2023 of students who started their studies and either graduated or dropped out of the program in this time interval.

We used the enrollment and exam results of the students. The grading scale is [1.0, 1.3, 1.7, 2.0, 2.3, 2.7, 3.0, 3.3, 3.7, 4.0, 5.0], where the best grade is 1.0, the worst is 4.0, and 5.0 means failing. Data exploration shows that marks are not distributed the same way in both universities: while the five most given marks

Table 2. Vectorial representation of a student at the end of Semester 1.

M01_1_N	M02_1_N	M03_1_N	M04_1_N	M05_1_N	M06_1_N
5.0	2.3	3.3	5.1	5.0	5.1

in courses from semesters 1 to 3 are 1.3, 1.7, 1.0, 2.0, and 5.0 in this order at T, they are in the order 1.0, 1.3, 1.7, 2.0 and 5.0 at B. Data exploration shows also that the time to graduate and the time to drop out varies between the two universities: 50% of the students need nine semesters or less to graduate at T, while they need 10 semesters or less at B; 50% of the students spend three semesters or less in the study program before they drop out at T while they spend five semesters or less at B. In both universities, as already written, most students study part-time: 50% of the students take four exams or less per semester.

Table 1 shows the number of students in our datasets at the end of the 1st, 2nd, and 3rd semesters in the two universities B and T according to their status "graduate (G)" or "dropout (D)". For example, there are 183 students at the end of semester 2 at B, 156 of whom dropped out, see rows Sum and D[1].

Row Sum of Table 1 shows that there are more students at T than at B. Table 1 shows also that data is imbalanced: only about 15% (27/183) of the students who are in their 2nd semester of studies have finally graduated (G) at B and 19% at T. Data is less imbalanced at T than at B, less imbalanced in semester 3 (S3) than in semester 2 (S2) than in semester 1 (S1).

3.2 Methodology

We predicted dropout or graduation after the 1st, 2nd, and 3rd semesters. Data exploration has shown that the bulk of the students drop out during their first three semesters. We conducted predictions considering students from each university separately (RQ1), and then merging the students from the two universities into one dataset (RQ2). This is possible because the study programs have the same courses and have the same handbooks.

Feature Set. As already written above, we used a set of courses as features to represent students. Our data exploration has shown that, though students must not follow the plan of the study handbook, most of them do so in their first semester of study. Further, as most of the students study part-time, few of them

[1] The discrepancy in the number of graduates at University B is attributed to inconsistencies found in the initial unprocessed data. We have the information about the graduation status for all the students, but not the corresponding grades for all the semesters of their degree courses. This is why we lost one student in the 2nd and 3rd semesters.

Table 3. Additional features to predict dropout at the end of the 2nd semester.

M01_2_N	M02_2_N	M03_2_N	...	M11_2_N	M12_2_N
1.7	2.3	3.3	...	5.1	1.7

enroll in all courses of a semester as planned in the handbook. This means that each course has missing values. Missing values were imputed with 5.1, close but different from 5.0 (fail); since all courses of the 1st, 2nd, and 3rd semesters are mandatory, a student must pass them all sooner or later to graduate, this was the reason to choose a value close to 5.0 (fail) as missing value; choosing a higher value than 5.1, like 6.0, could act as a penalty, what should not be done as many students study part-time, see Section 3.1. Such a representation is shown in Table 2. M01, M02, ... till M06 are the codes of the six mandatory modules of the 1st semester as described in the study handbook. The student represented in Table 2 failed M01 and M05 (M01_1_N and M05_1_N are 5.0), passed modules M02 and M03 with the mark 2.3 and 3.3 respectively, and did not enroll the modules M04 and M06 (M04_1_N and M06_1_N have the missing value 5.1). Six features thus represent the academic performance of a student in the first semester.

To predict dropout or success at the end of the second semester, we used all courses planned for the two first semesters in the study handbook, much in the same way as described above. However, we remembered the history by repeating the courses of the first semester. Table 3 shows the features that are added to the features of the first semester to predict dropout at the end of the second semester. The student enrolled again in M01 and passed it with the mark 1.7 (M01_2_N is 1.7). Because students cannot repeat modules that they have passed, the values M02_2_N and M03_2_N are the values of M02_1_N and M03_1_N in Table 2. This student enrolled in the 2nd-semester course M12 and passed with a mark of 1.7 but did not take the course M11 (M11_2_N is 5.1). When predicting dropout at the end of the 2nd semester, a student is thus represented by 18 features: the six features showing marks of the 1st semester, and 12 additional features showing marks (or missing values) he/she obtained in the 2nd semester for courses of the 1st and 2nd semesters according to the handbook.

To predict dropout or success at the end of the 3rd semester, we proceeded similarly to the 2nd semester. As the handbook foresees six mandatory courses in each of the first three semesters, the feature set to predict dropout or success at the end of the 3rd semester has 36 features.

Training, Testing, and Hyperparameter Tuning. Some authors select for the test set students who started the degree program most recently, reflecting the intended use of the prediction: models built with passed students are used to predict dropout for new students [27]. We could not use such an approach because our dataset is too small. Therefore, we performed a classical 5-fold cross-validation to train and test the models and report the mean results. While training the models, we performed again a (nested) five-cross validation for hyperparameter optimization using a grid search, following the same approach as [8].

Algorithms. We have used binary decision tree (DT) and logistic regression (LR). The algorithm DT builds a tree from the training data. The root of the tree contains the full training set. Then, features are selected recursively to divide the data into two subsets that are more homogeneous for the class to be predicted. To classify an element, a path is followed in the tree starting with the root till a leaf is reached. At each node, the decision is made according to the value that an element has for the given feature. The majority class of a leaf determines the prediction. We have used Python and the scikitlearn[2] library with the following grid search parameters: the maximal depth of the tree (range: 4 to 12) and the criterion to divide a node ('gini', 'entropy'). Other parameters such as the maximum number of leaf nodes and the complexity parameter (ccp_alpha) did not improve the balanced accuracy.

The logistic regression calculates optimal coefficients to all features using the training data; to classify an element, it performs a linear combination of the values of the features for this element using the coefficients and then applies the logistic function. We have used the solver ('liblinear', 'saga', 'newton-cg', 'lbfgs', 'sag') and the penalty ('l1', 'l2') as our grid search parameters.

Data Balancing. We have not performed any data balancing using a popular approach like SMOTE [11] because we have observed that, depending on how the synthetic data is generated, different courses are chosen to split the data with decision trees, see also [22] for a work that investigates how data balancing approaches might affect models' behavior. Instead, we have used the option class_weight = 'balanced' offered in scikitlearn. Each instance is weighted by the inverse proportion of the class frequency. Thus, the instances of the minority class have a higher weight than the ones of the majority class.

Model Evaluation. We have used the following metrics:
Accuracy (ACC): proportion of correct predictions.
Recall (REC), also called true positive rate (TPR): proportion of students who dropped out and are correctly predicted to drop out.
Specificity (SPE), also called true negative rate (TNR): proportion of students who graduated and are correctly predicted to graduate.
Balanced Accuracy (BAC): mean of REC (TPR) and SPE (TNR). As already mentioned, we considered this metric to be the most important as our data is imbalanced and used it for hyperparameter optimization.

4 Results and Discussion

We present the results and discuss our research questions in turn.

[2] https://scikit-learn.org

Table 4. Performance in % of the decision trees (DT) and logistic regression (LR) at the end of the 1st semester (S1), 2nd semester (S2), and third semester (S3) using the data of universities T and B and the metrics balanced accuracy (BAC), recall (REC), specificity (SPE), and accuracy (ACC). The values in bold are the highest for the metric.

	DT						LR					
	S1		S2		S3		S1		S2		S3	
	T	B	T	B	T	B	T	B	T	B	T	B
BAC	92.0	74.0	78.4	68.0	85.8	74.4	89.8	82.4	80.4	69.0	83.2	71.3
REC	91.0	98.0	96.8	92.5	77.8	88.8	90.8	89.8	80.8	87.9	80.6	82.6
SPE	93.4	50.0	60.0	43.7.	93.0	60.0	88.9	75.0	80.0	50.0	85.7	60.0
ACC	91.4	94.0	90.8	87.0	82.6	83.6	90.6	88.7	80.6	83.8	82.2	78.6

RQ1: What is the performance of the classifiers 'decision tree' and 'logistic regression' when predicting whether students are at risk of dropping out of the online degree? Which courses are important for the prediction? Table 4 presents the results of the prediction after the 1st semester (S1), 2nd semester (S2), and the 3rd semester (S3). In this part, we consider the prediction done with the separate data sets from the two universities, columns T and B. Looking at column DT, then S1, and the metric recall (REC), for example, we see that the algorithm DT can find 91.0% of the students who dropped out in University T and 98.0% in University B. We observe that the values for specificity (SPE) vary the most across columns. This could be due to the small number of students with the status "graduate" in the data. We also notice that the values of the metrics tend slightly to be higher in University T than in University B. This could be due to the bigger dataset of University T. Finally, none of the two algorithms clearly wins. Note, however, that DT gives better results than LR for S1 at University T. We have compared these values to the values obtained in other works that used local features. The recall values tend to be comparable to the values obtained by Wagner et al. in three face-to-face study programs using five different algorithms while the balanced accuracy values tend to be slightly lower [27]. We have also compared the values of Table 4 with the values obtained by Manrique et al. in two face-to-face study programs using four different algorithms [17]. The recall and accuracy values tend to be comparable to those of [17].

Like the results of Wagner et al. [27] but contrary to the results of Berens et al. [4] and Manrique et al. [17], the results are not better when the prediction is made for a higher semester. For example, balanced accuracy using LR drops from 89.8% at T in semester 1 to 83.2% in semester 3 and at B from 82.4% to 71.3%. This could be because the datasets get smaller, see Table 1.

Discussing the features and courses impacting the results. We propose to establish a list of courses that impact the most prediction separately for

Table 5. The seven most important features for decision tree DT and logistic regression LR for each of the semesters S1, S2, and S3 in University T.

DT-S1	LR-S1	DT-S2	LR-S2	DT-S3	LR-S3
M04_1_N	M01_1_N	M02_2_N	M02_2_N	M02_3_N	M16_3_N
M01_1_N	M05_1_N	M08_2_N	M07_2_N	M09_2_N	M06_3_N
M03_1_N		M07_2_N	M09_2_N	M09_3_N	M12_2_N
M05_1_N		M11_2_N	M02_1_N	M11_2_N	M12_3_N
M02_1_N		M04_2_N	M01_1_N	M05_1_N	M10_3_N
M01_1_N		M09_2_N	M06_2_N	M10_3_N	M02_2_N
		M07_2_N			M01_2_N

each university because merging the two datasets did not produce systematically better results as we will see below. Because there is no clear winner as written above, we use the results of both algorithms to establish such a list. For each model, we list the features in order of importance for the prediction. For the decision trees, the importance is given by the number of elements in the node. Presently, we consider nodes that contain at least 10% of the dataset. This limit should be discussed with our stakeholders. For the logistic regression, the order is given by the absolute value of the coefficients. We consider features with a coefficient bigger than 0.4 in absolute value; as above, this threshold has to be discussed with our stakeholders. For each model, we select the seven most important features in case the model uses more than seven important features. The number 7 has been chosen to not overload the overview with too many features. Table 5 shows the list of the most important features for each model built to predict dropout at University T. The modules M01, M02, ... till M06 are scheduled in the 1st semester in the study handbook, the modules M07, M08, ... till M12 in the 2nd semester, and the modules M13 till M18 in the 3rd semester. One notices features concerning enrollment or marks that students have in their second semester for courses scheduled in the 1st semester, for example, the features M02_2_N, and M06_2_N in column LR-S2. This reflects the fact that many students study part-time. Notice that a feature can appear several times in a decision tree, like M01_1_N in column DT-S1.

The classifiers use the history of the students, see for example features M02_1_N and M01_1_N in column LR-S2. One notices that most of the important features in semester 3, S3, concern courses scheduled in the first or second semester.

From Table 5 we extract the courses linked to the features and count for each course how many features are linked to it. For example, course M02 is linked to the six features: M02_1_N, M02_2_N, M02_2_N, M02_1_N, M02_3_N, and M02_2_N. As a summary, we propose to return to stakeholders the courses whose linked features are mentioned at least three times in the six models as the list of the courses that impact the prediction the most. For University T, this list contains four courses: M02 “Principles of Programming 1“ (6 linked

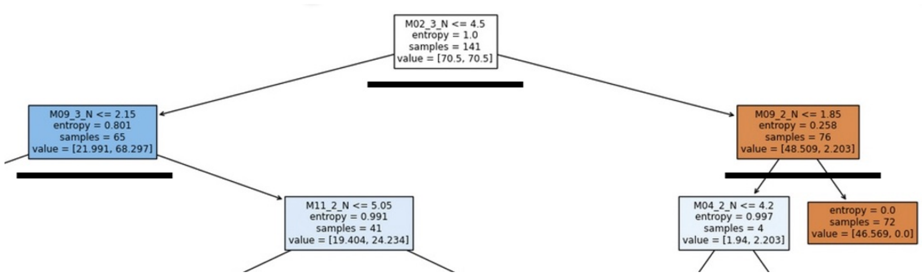

Fig. 1. A part of the decision tree associated with the prediction for semester 3 at University T. Nodes with important courses are underlined.

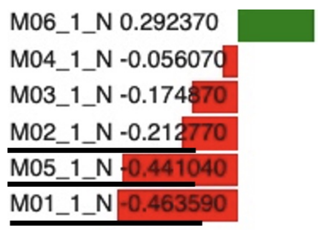

Fig. 2. The coefficients of the logistic regression model associated with the prediction for semester 1 at University T. Coefficients associated with important courses are underlined.

features), M01 “Introduction to Computer Science” (4), M09 “Human-Computer Interaction” (4), M05 “Media Design (3).

However, a preliminary discussion with stakeholders has shown that such a list with no context is not very helpful. We propose associating the three decision trees to the list, highlighting the parts containing the impactful courses, and also the coefficients of the features, highlighting those linked to the courses of the list. As an example, Figure 1 shows a part of the tree associated with the column DT -> S3 -> T of Table 4, the decision tree predicting dropout at the end of semester 3, and Figure 2 shows the coefficients for the column LR -> S1 -> T, the coefficients of the linear regression predicting dropout at the end of semester 1.

The negative coefficients in Figure 2 show that a big value of these features, in our case this means a bad mark, or fail or no mark, contributes to the prediction dropout. The other courses shown in Figure 2 are M03 “Communication, Leadership and Self-Management”, M04 “Principles of Mathematics”, and M06 “Computer Architecture and Operating Systems”.

RQ2: Does predicting dropout for each university separately give a better performance than merging the two academic datasets? Table 6 repeats the results of Table 4 and adds the results of the prediction with logistic regression when the

Table 6. Average Performance in % of the logistic regression at the end of the 1st (S1), 2nd (S2), and 3rd semester (S3) using the data of universities T and B and merging both (T+B).

	S1			S2			S3		
	T	B	T+B	T	B	T+B	T	B	T+B
BACC	89.8	82.4	69.1	80.4	69.0	79.2	83.2	71.3	84.0
REC	90.8	89.8	85.0	80.8	87.9	87.8	80.6	82.6	93.0
SPEC	88.9	75.0	53.3	80.0	50.0	70.6	85.7	60.0	75.0
ACC	90.6	88.7	81.8	80.6	83.8	84.8	82.2	78.6	89.0

two datasets are merged, see column T+B. One does not notice an overall improvement in the results, except partly in semester 3. Similar results have been obtained with decision trees. The results suggest that predicting 'dropout' for each university separately gives a better performance than predicting dropout with a merged dataset. It is interesting to note a parallel with the work of Wagner et al. [27]. The authors have used global features to predict dropout, which allows for merging data from different study programs. The model trained with a dataset after merging the data from the three study programs did not perform equally well when tested on data from the individual study program.

5 Conclusion, Limitations, and Future Work

In this work, we have used two interpretable algorithms, decision tree, and logistic regression, to predict whether students will drop out of their studies in an online study program offered by two universities. We make the prediction at the end of the 1st, 2nd, and 3rd semesters of study. The results of the first research question concerning the prediction are overall comparable to the results that others have obtained in face-to-face study programs. We observe that overall, none of the two algorithms give better results than the other. We propose to rank the features of all models by order of importance and extract from these ranked lists of features a list of courses that impact the predictions the most. This list is to be communicated to stakeholders. Further, we propose to supplement it with decision trees highlighting the relevant courses as well as the values of the coefficient of the features linked to the courses in the list. We hypothesize that this contextual information will help stakeholders to make sense of the results.

Our results also show that merging the data in one single dataset does not give better results than predicting dropout separately for each university, addressing our second research question.

Our work has several limitations. One limitation is that we have not checked whether the models work equally well for female and male students and thus are fair concerning gender. We have considered all students stored in the information system. However, in Germany, there are so-called “Scheinstudierenden”, students who do not have the intention of graduating when they register for university, see [21]. The presence of such students in our dataset could affect the results. Data exploration so far has not given reliable indicators to detect and exclude them. Finally, our research has used only two interpretable algorithms to predict students at risk of dropping out.

We have presented our findings and our approach to pinpoint important courses impacting the prediction the most to eight heads of programs and teachers of five different universities of the network. The stakeholders found the results understandable, and their feedback was positive. They saw the potential to help pay attention to critical courses as well as to shape their onboarding event at the start of the first semester differently; one suggested that students should be better aware of critical courses as they begin to study. Further research is needed. It should be investigated whether and to which extent the decision trees and the coefficients should be supplemented with explanations in natural language.

We plan to use other algorithms, in particular, XGBoost known to give good results in many situations, also with imbalanced datasets, see [10]. A challenge is to generate explanations. Although there are methods to explain the predictions of black-box models such as XGBoost [7,19], the work of Swamy et al. shows that these methods should be chosen with care [23]: comparing several approaches has revealed that the selected approach has a far greater impact on the features used to explain the prediction than the underlying data.

Acknowledgments. This study was funded by the German Ministry of Education and Research under the grant number 01PX21001A.

Disclosure of Interests. Agathe Merceron had no involvement with the workshop's handling of this article in order to avoid a conflict with her role as co-organizer of the workshop. All other authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arnold, K., Pistilli, M.: Course signals at purdue: using learning analytics to increase student success. In: Proceedings of the 2nd International Conference on Learning Analytics & Knowledge (LAK). pp. 267–270. Association for Computing Machinery, New York, NY, USA (2012). https://doi.org/10.1145/2330601.2330666
2. Aulck, L., Nambi, D., Velagapudi, N., Blumenstock, J., West, J.: Mining University Registrar Records to Predict First-Year Undergraduate Attrition. In: 12th International Conference on Educational Data Mining (EDM). pp. 9–18. International Educational Data Mining Society, Montreal, Canada (2019), https://eric.ed.gov/?id=ED599235
3. Baker, S., Esbenshade, L., Vitale, J., Karumbaiah, S.: Using demographic data as predictor variables: a questionable choice. Journal of Educational Data Mining **15**(2), 22–52 (2023). https://doi.org/10.5281/zenodo.7702628, https://jedm.educationaldatamining.org/index.php/JEDM/article/view/619

4. Berens, J., Schneider, K., Gortz, S., Oster, S., Burghoff, J.: Early Detection of Students at Risk - Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods. Journal of Educational Data Mining **11**(3), 1–41 (2019). https://doi.org/10.5281/zenodo.3594771
5. Cai, C., Fleischhacker, A.: Structural neural networks meet piecewise exponential models for interpretable college dropout prediction. Journal of Educational Data Mining **16**(1), 279–302 (2024). https://doi.org/10.5281/zenodo.11236277
6. Cannistra, M., Masci, C., Ieva, F., Agasti, T., Paganoni, A.M.: Early-predicting dropout of university students: An application of innovative multilevel machine learning and statistical techniques. Studies in Higher Education **47**(9), 1935–1956 (2022). https://doi.org/10.1080/03075079.2021.2018415
7. Cohausz, L.: When probabilities are not enough - a framework for causal explanations of student success models. Journal of Educational Data Mining **14**(3), 52—-75 (2022). https://doi.org/10.5281/zenodo.7304800
8. Cohausz, L., Tschalzev, A., Bartelt, C., Stuckenschmidt, H.: Investigating the importance of demographic features for edm-predictions. In: Feng, M., KÃ¤ser, T., Talukdar, P. (eds.) Proceedings of the 16th International Conference on Educational Data Mining. pp. 125–136. International Educational Data Mining Society, Bengaluru, India (July 2023). https://doi.org/10.5281/zenodo.8115647
9. Dekker, G., Pechenizkiy, M., Vleeshouwers, J.: Predicting students drop out: A case study. In: Barnes, T., Desmarais, M., Romero, C., Ventura, S. (eds.) Proceedings of the 2nd International Conference on Educational Data Mining. pp. 41–50. International Educational Data Mining Society, Cordoba, Spain (July 2009). https://doi.org/10.5281/zenodo.8115647
10. Gardner, J., Popović, Z., Schmidt, L.: Subgroup robustness grows on trees: an empirical baseline investigation. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. p. 127–203. NIPS '22, Curran Associates Inc., Red Hook, NY, USA (2022)
11. Han, H., Wang, W.Y., Mao, B.H.: Borderline-smote: a new over-sampling method in imbalanced data sets learning. In: Proceedings of the 2005 International Conference on Advances in Intelligent Computing - Volume Part I. p. 878–887. ICIC'05, Springer-Verlag, Berlin, Heidelberg (2005). https://doi.org/10.1007/11538059_91, https://doi.org/10.1007/11538059_91
12. Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., König, R., Richter, J., Woisch, A.: Motive und ursachen des studienabbruch an baden-wurttembergischen hochschulen und beruflicher verbleib der studienabbrecherinnen und studienabbrecher. Tech. rep., DZHW (2017)
13. Heublein, U., Hutzsch, C., Schmelzer, R.: Die Entwicklung der Studienabbruchquoten in Deutschland. (DZHW Brief 05|2022). Hannover: DZHW (2022), https://doi.org/10.34878/2022.05.dzhw_brief
14. hochschulkompass Stiftung zur Förderung der Hochschulrektorenkonferenz: Studium - fernstudium, https://www.hochschulkompass.de/studium/rund-ums-studieren/studienformen/fernstudium.html
15. Jayaprakash, S., Moody, E., Lauría, E., Regan, J., , Baron, J.: Early alert of academically at-risk students: An open source analytics initiative. Journal of Learning Analytics **1**(1), 6–47 (2014), https://learning-analytics.info/index.php/JLA/article/view/3249
16. Kemper, L., Vorhoff, G., Wigger, B.U.: Predicting Student Dropout: A Machine Learning Approach. European Journal of Higher Education **10**(1), 28–47 (2020). https://doi.org/10.1080/21568235.2020.1718520

17. Manrique, R., Nunes, B.P., Marino, O., Casanova, M.A., Nurmikko-Fuller, T.: An Analysis of Student Representation, Representative Features and Classification Algorithms to Predict Degree Dropout. In: Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK). pp. 401–410. Association for Computing Machinery, New York, NY, USA (2019). `https://doi.org/10.1145/3303772.3303800`
18. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence **267**, 1–38 (February 2019). `https://doi.org/https://doi.org/10.1016/j.artint.2018.07.007`
19. Molnar, C.: Interpretable machine learning (2024), https://christophm.github.io/interpretable-ml-book/
20. de Quincey, E., Briggs, C., Kyriacou, T., Waller, R.: Student centred design of a learning analytics system. In: Proceedings of the 9th International Conference on Learning Analytics & Knowledge. p. 353–362. LAK19, Association for Computing Machinery, New York, NY, USA (2019). `https://doi.org/10.1145/3303772.3303793`
21. Schneider, K., Berens, J., Görtz, S.: Maschinelle fruherkennung abbruchgefärdeter studierender und wirksamkeit niedrigschwelliger intervention. In: Neugebauer, M., Daniel, H., Wolter, A. (eds.) Studienerfolg und Studienabbruch, pp. 367–391. Springer VS (2021)
22. Stando, A., Cavus, M., Biecek, P.: The effect of balancing methods on model behavior in imbalanced classification problems (2023), arXiv:2307.00157
23. Swamy, V., Radmehr, B., Krco, N., Marras, M., KÃ¤ser, T.: Evaluating the explainers: Black-box explainable machine learning for student success prediction in MOOCs. In: Mitrovic, A., Bosch, N. (eds.) Proceedings of the 15th International Conference on Educational Data Mining. pp. 98–109. International Educational Data Mining Society, Durham, United Kingdom (July 2022). `https://doi.org/10.5281/zenodo.6852964`
24. Tinto, V.: Dropout from higher education: A theoretical synthesis of recent research. Review of Educational Research **45** (1975)
25. Wagner, K., Hilliger, I., Merceron, A., Sauer, P.: Eliciting Students' Needs and Concerns about a Novel Course Enrollment Support System. In: Companion Proceedings of the 11th International Conference on Learning Analytics & Knowledge (LAK). pp. 294–304. Online (2021), https://www.solaresearch.org/core/lak21-companion-proceedings/
26. Wagner, K., Merceron, A., Sauer, P., Pinkwart, N.: A course recommender system built on success to support students at risk in higher education. Journal of Educational Data Mining **16**(1), 330–364 (Jun 2024). `https://doi.org/10.5281/zenodo.11384083`
27. Wagner, K., Volkening, H., Basyigit, S., Merceron, A., Sauer, P., Pinkwart, N.: Which Approach Best Predicts Dropouts in Higher Education?:. In: Proceedings of the 15th International Conference on Computer Supported Education (CSEDU). pp. 15–26. INSTICC, SciTePress, Prague, Czech Republic (2023). `https://doi.org/10.5220/0011838100003470`
28. Wöhrle, J., Karolyi, H., Weidlich, J., Blanc, B., Woick, A.: Wechselwirkungen zwischen digitalisierter hochschulbildung und studienerfolg. In: Wissenschaftskonferenz: Studienerfolg durch Digitalisierung? FernUniversität in Hagen (2024), https://www.fernuni-hagen.de/forschung/schwerpunkte/catalpa/aktuelles-termine/aktuelles/wissenschaftskonferenz-lamass.shtml