Computer Science and Information Systems 18(3):957–978 https://doi.org/10.2298/CSIS200920053R

Predicting Dropout in Online Learning Environments

Sandro Radovanović[1], Boris Delibašić[1], and Milija Suknović[1]

[1] University of Belgrade - Faculty of Organizational Sciences
11000 Belgrade, Serbia
{sandro.radovanovic, boris.delibasic, milija.suknovic}@fon.rs

Abstract. Online learning environments became popular in recent years. Due to high attrition rates, the problem of student dropouts became of immense importance for course designers, and course makers. In this paper, we utilized lasso and ridge logistic regression to create a prediction model for dropout on the Open University database. We investigated how early dropout can be predicted, and why dropouts occur. To answer the first question, we created models for eight different time frames, ranging from the beginning of the course to the mid-term. There are two results based on two definitions of dropout. Results show that at the beginning AUC of the prediction model is 0.549 and 0.661 and rises to 0.681 and 0.869 at mid-term. By analyzing logistic regression coefficients, we showed that at the beginning of the course demographic features of the student and course description features are the most important variables for dropout prediction, while later student activity gains more importance.

Keywords: Education Data Mining, Learning Analytics, Dropout prediction, Lasso, and Ridge Logistic Regression.

1. Introduction

Over the past few decades, education systems had trouble responding to market requirements. Namely, skills and knowledge needed for the industry include technologies that are just developed, thus leaving educational systems no time for full curriculum and syllabus development which could blend into existing study programs. European Commission recognized the problem and developed a term called short cycles of education that are intended for people who want to learn a specific subject, without studying the whole study program [5]. This way students or interested parties can participate and obtain needed knowledge for the task at hand. However, short cycles of education required in house training or supervision of the student which discouraged many of students or professionals. For example, students had to be physically present at the teaching center or they had to study only during the classes. The full bloom of short cycles of education is noted with the development of Massive Open Online Courses (MOOCs) which allows access to learning materials from worldwide renowned Universities and professors on a variety of subjects [29]. Using MOOC platforms one can tailor a learning path to its preferences. Additionally, the learning path can be achieved at any course order, at any pace, without being present and often free of charge [8]. These benefits attracted a lot of students and professionals.

However, newly founded flexibility of learning which includes a diversity of subjects and ease of access to learning materials triggers new problems not observed in

traditional learning environments. The major problem in MOOCs is a low percentage of students finishing courses. This phenomenon is called dropout and is defined as a student that unenrolled from course materials before the formal end of the course [44] or a student failing to obtain a passing grade for the course [45]. Although some of the students enroll in the online course just to obtain learning materials, some students interacted with a learning environment, i.e. listened to the lectures, read additional materials, tried quizzes and assignments, and did not obtain enough points to obtain a certificate of accomplishment. Reasons for failing the course can be insufficient background knowledge, lack of time, course design, or one felt discouraged, frustrated, or bored [17].

A lot of research efforts by academia and course providers are invested in answering the question of how and why students dropout. This area of research is studied under a wider discipline called Learning Analytics or Educational Data Mining [35]. Application of data mining or machine learning to education domain is needed [36] because the lack of "negative samples", i.e. due to the fact that majority of the students are considered as a dropout and that there are a large number of students which in classical statistical analysis results in significant impacts even if it is not. Another issue is a large volume of unstructured data. Although unstructured data is potentially very informative, one must put them into a structure and derive attributes that describe student behavior in a learning environment. The third problem is data variance which is the result of self-paced learning. Namely, students can have many different learning styles which all result in a certificate of accomplishment. One student can interact with the learning environment on a regular basis, do assignments and quizzes, while others can just take assignments, another can just listen to lectures, or some can download learning materials and listen to them on their computer. Each student may finish the course, but they all had different behavior leading to it. This poses a problem to traditional statistical testing so machine learning methods are considered as an appropriate approach. Also, the point of interest that classical statistical analysis fails to address is error analysis. Namely, the course designer wants an analytical model that has a low number of false-negative students, i.e. course designer wants to identify everyone who will fail to pass an exam and contact them as early as possible. This can come with a cost of false alarms (students who are identified as students who are going to fail an exam, but they are going to pass the exam).

In this paper, we used the Open University Learning Analytics dataset [28] to develop models for student dropout prediction. Open University dataset contains 22 courses with different behavior of the student, i.e. interaction with the learning environment (watching videos, reading materials, etc.), scores on quizzes and assignments, and historical enrollments. Due to multiple definitions of dropout, we use two experimental setups with two definitions of dropout (both will be called under umbrella term dropout), one presenting prediction model for students who fail to pass an exam or unenroll from the course (i.e. unenrolled from the course or student did not achieved enough points for the certificate), and the other presenting prediction model for students who unenrolled (i.e. unenrolled from the course). The goal of this paper is to identify how early we can predict dropout. Therefore, besides having two prediction models we will try to predict dropout as early as possible. The dropout models are produced with logistic regression as an algorithm because it provides interpretable models and because it is very suitable to work with datasets with a lot of attributes, and because it tries to fit the distribution of classes (dropouts and successful candidates) in

the predictive model respecting the conditional distributions of all attributes w.r.t. the class attribute. To add, coefficients of logistic regression can be interpreted in terms of the logarithm of odds of dropout which can be seen as either a positive or negative influence on dropout.

The contributions of the paper are solving the problem based on two definitions of dropout. Namely, the majority of the papers utilize one definition of the problem which is easier to solve (student will fail to pass the exam or withdraw from the course). Because of that, we developed two predictive models. One, where dropout is defined as a student who fails to pass an exam [36] and another, where a dropout is defined as a student who will withdraw from or fail to pass the exam [43, 40]. Besides using two definitions of the dropout, we created and evaluate logistic regression models in eight different time frames, ranging from the beginning of the course up to the mid-term of the course. Therefore, we provide an answer to how early can we predict a dropout. An additional contribution of the paper is the utilization of the aggregation functions which are not commonly used in learning analytics. In order to gain better results, we used recency [10] and variability seeking index [12] which have shown importance in marketing and sports, respectively. In addition, we utilize counterfactual examples [42] that can aid decision-makers in helping the student by providing causal reasoning on how to reach a positive outcome. Predictive performance is done using the area under the curve (AUC) and area under the precision-recall curve (AUPRC), which can be found in the papers. However, we provide cumulative gain charts and lift curves which are useful for decision making of the predictive model. Finally, we analyzed coefficients of logistic regression to give an insight into why students are becoming dropouts. Since there are eight different time frames we interpret coefficients of the logistic regression and give possible answers to why dropout occurs for a different time period of the course. This finding can be used for course makers and course designers for dropout prevention strategies.

The remainder of the paper is organized as follows. In Section 2 we present a review of the literature. In Section 3 we present methodology. We will present data used in this research, followed by the experimental setup and evaluation of the predictive model. In Section 4 we provide results and interpretation of results, while in Section 5 we conclude the paper.

2. Literature review

From a historical point of view, the first MOOC called “Connectivism and Connective Knowledge” was created by George Siemens and Stephen Downs in 2008 which attracted over 2,000 students who participated free of charge [14]. Today, MOOCs environments such as Coursera, EdX, or Udacity have courses with over 1,000,000 enrolled students coming from over 190 countries [37].

Due to a proliferation of MOOCs in past years and the fact that a minority of students complete course, researchers from the technical field such as statistical analysis, data mining, and machine learning alongside with domain experts in fields of pedagogy, education, and organization tried to tackle the problem of dropout prediction and prevention. The first, main challenge, was the ill definition of the term dropout. The most common, term dropout (or stopout) is defined as a moment when a student

unenrolled from the course. From that point, the student does not have access to learning materials anymore. However, many students do not unenroll from the course, but their activity is very low if existing at all. Therefore, the term dropout should be redefined to the last event student participated in, such as a quiz assignment [40] or watching a video [2]. We can define them as students who stopped participating in the course. We can ask ourselves, what about students who participated in the course and yet failed to pass the exam? Are they also dropouts? In some sense, they can be considered as dropouts. They had trouble keeping up with course materials and they needed help. Having in mind that these systems are created to help students gain knowledge and skills needed for tomorrow, one can try to identify them in advance and help them, i.e. give more time for assignments, or provide additional readings. Therefore, many researchers defined dropout as a situation when a student does not earn a certificate of accomplishment within a course [25, 20, 9, 32].

To the best of our knowledge models for predicting fail on the online courses is set as a binary classification task for fixed time periods. Juang et al. [25] used only the performance of the student on the first-week assignment. In their example, only that information was enough to recognize which students will receive a certificate of accomplishment with distinction compared to students who received a certificate of accomplishment with AUC 0.947, and also between students with a certificate of accomplishment and students who failed to pass the exam with AUC 0.851. A similar application can be found in [33, 26]. Namely, student activities on quizzes and assignments are used to predict performance on the final exam. An approach that was used is based on matrix factorization where latent features that describes student cohort and interpret their importance to pass exam with three points of predictions, one at the beginning of the course, one at the mid-point, and final one, a week before the exam. It has been shown that performance increases as more information are added, i.e. more data is available. Namely, predictions are worst at the beginning of the course and increases as more information about student interaction and behavior is added.

However, students do have more activities during the class which can be used for the prediction model. In MOOCs, course providers often have clickstream data, which is considered as an unstructured data set. One can extract features that describe student interaction with the learning environment. For example, interaction with video learning materials, activity on the discussion forum, time spent on a specific page is used. In paper [40] logistic regression with several groups of attributes is used. One group of attributes are attributes that correspond to submission and problem solve such as the number of submissions, a number of distinct problems attempted, a distinct number of correct solutions, the average time needed for submission, etc. Another group is regarded as interaction with other students such as number of forum posts, number of forum responses, number of wiki edits, the total number of collaboration, or time spent of forum and wiki. These features are used for predicting whether a student will withdraw in the fifth week from the reference week (i.e. predicting one month in advance). Similarly, in paper [9] latent Dirichlet allocation is used for behavioral trend identification based on problem sets answers, interaction with questions, videos, forum, etc. It has been shown, as in previous researches, that more information about student performance provides better predictive performance (the longer the course lasts, the model is better at identifying dropout).

Analysis of dropouts on the dataset used in this paper has been already made. Prediction using time series is available in the paper [18]. Namely, student engagement

in a virtual learning environment is transformed into time series data which are further classified using time series forest algorithm. The results that are obtained are underperforming at the beginning of the series, i.e. beginning of the course, but improves as more data is available. The role of demographic data is presented on paper [34]. Namely, decision trees are used to predict failure on the exam [13]. This model can be applied before the course starts and it can achieve accuracy between 66% and 83%. One can also find framework Ouroboros [22, 23] that is demonstrated on this data. Finally, one can find the application of Naïve Bayes and Decision trees for course success prediction [3].

Compared to other approaches we will utilize every source of student interaction with the learning environment including demographic data, registration data, video interaction, quiz attempts, and assignments attempts. We will utilize a logistic regression model with lasso and ridge regularization. The reason for this is the fact that coefficients of logistic regression can be interpreted in terms of logarithms of odds of dropout, which will allow us to interpret the influences of each attribute to dropout. Next, we will have two experimental setups regarding the definition of the dropout. In this paper, we, therefore, adopted two definitions that are common in the literature. Further, we used multiple aggregation functions to summarize and describe student behavior which is not used in learning analytics at all, such as the recency of the event and variability seeking index. Finally, we utilized a logistic regression model that allows inspection of the coefficients. Based on the coefficients we can analyze the driving factors of a dropout.

It is noted that as a measure of performance most often Area Under the Receiver Operating Characteristics Curve (AUC) measure is used. AUC measures the probability that a classifier can discriminate between two randomly chosen data points, from which one is a positive outcome, and another negative is negative [1]. The reason why it is commonly used is that it is decision threshold independent, meaning that decision on what confidence or probability threshold predictive model will predict that student will fail an exam is omitted. Besides using AUC as a measure, we will use the area under the precision-recall curve (AUPRC) since it is more appropriate for class imbalance classification problems such as this one.

3. Data and Methodology

Data and Methodology section consists of an explanation of data and feature extraction from the database of Open University Learning Analytics Dataset [28]. Obtained data will be fitted into logistic regression. After an explanation of logistic regression, the whole experimental setup will be provided.

3.1. Data

Open University provided the database for learning analytics [28]. Data contain learning environment interaction, alongside with demographic data, enrollment data, etc. The Open University offers several hundred modules (subjects) from which every single one can be part of a university program or offered as a stand-alone course. Because

of that, it suffers from similar problems as MOOCs, i.e. dropout. Namely, a lot of students enroll in a specific subject, but due to various reasons did not finish or unenrolled from the course. However, Open University generates a better completion rate mainly because courses are offered for credit and the length of the course is around 9 months [22].

The database provided for analytics is anonymized and organized in a normalized manner containing seven tables presented in **Fig. 1**.

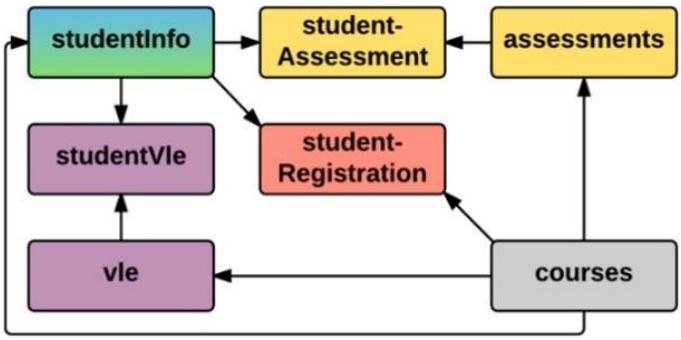

Fig. 1. Dataset structure [28]

In total 32,593 students registered to 22 courses. Information about the students is stored in table *studentInfo*, while information about the course is stored in table *courses*. During the course, the student had multiple assessments. Data about points achieved on the assessment is stored in table *student-Assessment*, while basic assessment information is stored in table *assessments*. There are 173,912 student assessment records. Finally, the student interacted with a virtual learning environment. There are 10,655,280 records of interaction and they are stored in table *studentVle*. Static information about the virtual learning environment is stored in table *vle*. In table *student-Registration* one can find information about registration of the student to the course and there is an indicator of the performance of the student on the course, i.e. withdraw, fail, passed, and distinction which is used for prediction.

Besides taking student demographic information (gender, region, highest education, IMD band, age, and disability), we generated aggregations (sum, count, mean, min, max, median, standard deviation, recency [976] and variability seeking index [976]) for student assessments and interaction with the learning environment. Two aggregation functions that are not common in many applications have been introduced. Namely, the recency of the event has shown to be of great importance in marketing [10]. Idea is to give more importance to events that occurred more recently. In other words, it gives decay to events that occurred long in the past. Variability seeking index is used for aggregation of categorical data, where difference compared to the previous event is calculated. If a student, i.e. changed the grade on the assessment then this deviance from the previous event should be accounted for. Variability seeking index has shown good predictive performance in sports [12].

Recency is calculated using the following formula (1).

$$recency = \sum_{i=1}^{m} s_i * 2^{-\left(\frac{x_i}{halflife}\right)} \quad (1)$$

where $halflife$ present interval for which effect of an attribute should be equal to 0.5 and x_i value to be inserted into the formula (i.e. days passed from the quiz). Since each student can have m events (quizzes or assignments), obtained recency scores are multiplied with the obtained score s_i and summed. This allows exponential decay of the effect of the obtained scores on the quiz or assignment. Half-life is always set to the half of the interval being predicted. For example, if we predict dropout based on the first-month activity, the half-life is equal to 15 days. In terms of educational data mining, this can be interpreted as forgetting term. More specifically, the effects of the previous quizzes and assignments are of less importance compared to the most recent ones.

Variability seeking index is an aggregation measure that calculates the trend of the scores obtained by the student. Although it does not satisfy all properties of the aggregation function (the result depends on the ordering of the data), this function allows identification of the subtle changes in the behavior of the student regarding the property one wants to analyze (i.e. activity on the learning environment, scores on the quizzes and assignments). It is calculated using formula (2).

$$vsi = \sum_{i=2}^{m} (s_i - s_{i-1}) \quad (2)$$

where s_i present the score obtained on the quiz or assignment in the time stamp i. A positive value will indicate an increase in the score values, or a positive trend in scores, while a negative value indicates a negative trend in the score values.

More specifically, an aggregated column from student assessment is a score on the assessment, point obtained from the assessment, and days submitted prior to the deadline. Aggregation is done on the student level and for each assessment type. For interaction with the learning environment number of clicks on the learning materials is aggregated on student level and activity level. In total, for each experiment, we extracted 522 features that describe student behavior.

3.2. Logistic regression

Logistic regression is one of the most popular machine learning algorithm with applications in various fields. It is commonly used in the educational domain, for dropout predictions [40, 25, 20, 23]. The main reason for the usage of logistic regression is the interpretability of the model. Namely, coefficients of the logistic regression model can be interpreted in terms of the odds ratio, and consequently in terms of probability. This property is important, especially for social science applications where each decision needs to be explained.

Logistic regression can be defined as a classifier that models the probability of dependent binary features *y* given a set of independent features *X* [19]. The model is defined as presented in the formula (3).

$$\log\left(\frac{p}{1-p}\right) = \theta_0 + \theta_1 x_1 + \cdots + \theta_k x_k. \quad (3)$$

where p represents the probability that dropout is going to occur ($y = 1$). Values of θ represent weights associated with independent features X. One wants to find the best values of θ such that the model provides the lowest possible error. Error is defined through loss function, called logistic loss (presented in the formula (4)), which has to be minimized.

$$\min \mathrm{L}(y, \hat{y}) = -\frac{1}{n}\sum_{i=1}^{n}(y_i \log(\hat{y}_i) + (1-y)\log(1-\hat{y}_i)). \quad (4)$$

One of the problems is that using logistic loss function one can overfit models (learn data at hand, without the power of generalization on the new examples) when working with a large number of attributes. Therefore, extensions of logistic regression have been developed to deal with the problem of overfitting in such situations. One can extend the logistic loss function to regularize the process of learning coefficients. With regularization, one intentionally makes a greater loss with a purpose to create a model that can generalize to new examples [24]. Lasso regularization adds L1 norm in $\mathrm{L}(y, \hat{y})$. L1 norm has the effect that coefficients of logistic regression are forced to zero, i.e. lowers the number of features needed to explain the problem at hand. This way complexity of the problem is reduced. Ridge regularization adds L2 norm which forces coefficients of logistic regression to be lower in general. Both regularization terms are used to prevent the model to explain random noise or error. However, regularization terms introduce hyper-parameter λ which needs to be optimized [41]. In this paper, we utilized inner 10-fold cross-validation to find the best λ that maximizes the AUC measure.

3.3. Experimental Setup

The goal of the paper is to answer two research questions. First, we would like to know how early we can predict whether the student will pass the exam and, second, we want to provide a discussion on what drives the student to fail an exam. In order to answer the first research question, we trained and tested logistic regression models in eight different time periods. The first model is created on day 0 of the course, as seen in [34]. In that period student does not have any assignment interaction. However, a predictive model can be created using demographic features and interaction with learning materials (since some of them are available before the course starts). The next predictive model is created after the first week (seven days) of course. This setup is common in the online course [25]. At this point, student generates data, i.e. interaction with learning materials (videos, readings, etc.) and prediction can already be made. The following time periods are after one month (30 days), 45 days, 60 days, 90 days, and 120 days (approximate middle of the course length). It is expected that the performance of the model will improve as more data about the interaction with the learning environment is available.

In order to answer the second research question, we utilized logistic regression and interpreted the coefficients of the logistic regression. More specifically, used lasso and ridge logistic regression and interpretation of coefficients was performed on the best

performing model. Coefficients of the logistic regression can be interpreted in terms of odds ratio and probability of dropout that can be found useful for course designing and decision-making. In addition, we utilize counterfactual examples. This powerful causal explanation finds the most related input attributes that lead to different outcomes [42].

Due to the fact that dropout has multiple definitions, we adopted two definitions which both present problem for any learning system as explained previously. Having in mind that we have eight different time frames of prediction we will have 16 experiments.

In order to have valid results students that unenrolled before the observed time period are dropped from the dataset. General information about a number of examples and the average dropout rate is presented in Table 1. As we can observe, a number of rows are exactly the same in both definitions of dropout. However, the percentage of the dropout is at the beginning two times greater if students that failed the course are included, increasing up to four times greater at the mid-term of the course (experimental setup where the model is created and evaluated after 120 days).

Models are evaluated using AUC because it is commonly used in educational data mining applications and specifically for the problem at hand. AUC can be interpreted as the probability that a random student who will fail to pass the exam has a greater probability that he/she will fail the exam than a random student who will pass an exam. This measure of evaluation is decision threshold independent, meaning that it is calculated for every possible combination of thresholds in data. A random classifier would have an AUC value of 0.5, while the perfect classifier would have an AUC value equal to 1 [11]. We also provide area under the precision-recall curve (AUPRC) which is also a common measure of classification model performance. It is interpreted as how many times a model is better compared to the default model. Values range from 0 to 1, where 1 is the value of the perfect classifier and the random classifier should have an average dropout rate in data at hand. Due to the fact that the model is evaluated using 10-fold cross-validation average value with the standard deviation will be presented. Additionally, we will present the lift curve and cumulative gain curve. Those model visualizations can be used for dropout prevention campaign definition and decision making. Namely, the lift curve presents a gain of a predictive model compared to using no model at all. On the x-axis percentage of students is presented, while on the y-axis present gain (ratio of the percentage of dropout students and the total number of students contacted) obtained using the predictive model. Value 1 on the y-axis presents a situation when the predictive model does not contribute to problem-solving, i.e. predictive model has no gain, while higher values improve the decision-making process (contact strategy). Having this in mind, the value of lift equal to 2 can be interpreted that the predictive model is two times better compared to using no model at all. The cumulative gain curve presents a comparison between dropout students and the total number of students. On the x-axis percentage of contacted students is presented, and on the y-axis percentage of dropout students are presented. If the gain curve is higher than the diagonal line, then the predictive model is usable. Namely, by contacting some percent of the students, we will be able to identify a higher percentage of dropout students.

Hyper-parameter λ for Lasso and Ridge regression was found using grid search with inner 10-fold cross validation.

Table 1. Dataset information

Dropout definition	Experimental setup	Number of rows	% of dropout
Withdraw	0 days	29,496	23.96%
	7 days	29,178	23.13%
	15 days	28,115	20.22%
	30 days	27,446	18.34%
	45 days	26,921	16.69%
	60 days	26,361	14.92%
	90 days	25,562	12.26%
	120 days	24,777	9.48%
Fail or Withdraw	0 days	29,496	47.84%
	7 days	29,178	47.27%
	15 days	28,115	45.28%
	30 days	27,446	43.99%
	45 days	26,921	42.85%
	60 days	26,361	41.64%
	90 days	25,562	39.81%
	120 days	24,777	37.91%

4. Results

After learning the model for both definitions of dropout following results are obtained. In Table 2 we present results on the withdrawal definition of dropout. One can observe that the performance of lasso logistic regression is better compared to ridge logistic regression for every experimental setup. Also, performance improves as more information about student behavior and interaction is available.

Table 2. Performance of logistic regression models on withdrawing students

Experimental setup	Lasso		Ridge	
	AUC	AUPRC	AUC	AUPRC
0 days	0.549 +/- 0.092	0.300 +/- 0.050	0.531 +/- 0.086	0.290 +/- 0.049
7 days	0.542 +/- 0.099	0.296 +/- 0.053	0.519 +/- 0.093	0.279 +/- 0.052
15 days	0.593 +/- 0.126	0.232 +/- 0.053	0.558 +/- 0.115	0.208 +/- 0.048
30 days	0.583 +/- 0.121	0.248 +/- 0.066	0.515 +/- 0.127	0.203 +/- 0.052
45 days	0.607 +/- 0.127	0.247 +/- 0.059	0.566 +/- 0.131	0.196 +/- 0.053
60 days	0.618 +/- 0.089	0.223 +/- 0.042	0.519 +/- 0.134	0.187 +/- 0.055
90 days	0.623 +/- 0.099	0.185 +/- 0.040	0.542 +/- 0.133	0.162 +/- 0.048
120 days	0.681 +/- 0.059	0.162 +/- 0.025	0.569 +/- 0.142	0.131 +/- 0.040

Initial model, i.e. at the beginning of the course, have trouble distinguish between dropouts and non-dropouts with AUC 0.549. This value of AUC means that model is just better than a random model. But, after the interaction of the student with learning materials and the learning environment model captures the withdrawal behavior and manages to discriminate between dropouts and non-dropouts. In the middle of the course (model created after 120 days), AUC is 0.681. A similar conclusion can be made based on AUPRC values. Namely, the initial model has a value 0.300 while the default

model should have 0.2396 (percentage of dropouts available in data). Based on these values, the predictive model is better by ~25% compared to a random model at the beginning of the course and ~71% after 120 days. This means that some features make a difference between dropout and non-dropout students.

The second definition of dropout students considers failing on the course and withdraw of the student as a dropout. To some extent, this definition is easier for prediction since some of the events, i.e. quizzes and assignments, directly influence the final grade. Performance in terms of AUC and AUPRC for this definition of dropout is presented in Table 3.

Table 3. Performance of logistic regression models on withdrawing and fail students

Experimental setup	Lasso		Ridge	
	AUC	AUPRC	AUC	AUPRC
0 days	0.661 +/- 0.054	0.635 +/- 0.056	0.651 +/- 0.054	0.628 +/- 0.054
7 days	0.669 +/- 0.062	0.644 +/- 0.061	0.660 +/- 0.059	0.637 +/- 0.059
15 days	0.673 +/- 0.074	0.633 +/- 0.077	0.663 +/- 0.068	0.624 +/- 0.070
30 days	0.693 +/- 0.089	0.646 +/- 0.097	0.707 +/- 0.081	0.658 +/- 0.089
45 days	0.738 +/- 0.087	0.683 +/- 0.096	0.739 +/- 0.084	0.686 +/- 0.090
60 days	0.788 +/- 0.052	0.753 +/- 0.063	0.791 +/- 0.045	0.756 +/- 0.054
90 days	0.820 +/- 0.036	0.791 +/- 0.041	0.817 +/- 0.038	0.785 +/- 0.045
120 days	0.869 +/- 0.025	0.841 +/- 0.030	0.864 +/- 0.033	0.833 +/- 0.040

Considering this definition of a dropout we obtained better predictive performance. Namely, AUC is at the beginning of the course 0.661 and improves up to 0.869. As in previous results, lasso logistic regression is mostly better compared to ridge logistic regression. Based on AUPRC values we can conclude that the predictive model is better than a random model for ~33% at the beginning of the course and ~122% after 120 days.

In order to answer the question of how early can we predict one must ask a question of what good enough performance of the model is? There are no formal guidelines for evaluation AUC and AUPRC values, i.e. what is considered as good performance. All of the models are useful. More specifically, they are better than uninformed decision making. Using the rule of thumb, AUC of 0.700 is considered as a good model. However, this value can be misleading if a class imbalance is present and AUPRC is recommended [11]. Our AUPRC suggests that our model is even at the beginning of the course better ~25% for withdrawing students and ~33% for withdrawing and failed students than a random model. We can say that models are usable, i.e. can be used for course design and decision making. For example, the model can be used for a mass campaign for the prevention of the dropout. Course designers could move the deadline, provide additional readings, or involve more details to students that are more prone to be a dropout.

5. Discussion

After presenting and discussing the results of the predictive model in terms of predictive performance, we present a discussion of the application of the predictive model. First,

we need to answer the question of whom to contact. For that purpose, we can use the lift curve and cumulative gain curve presented in **Fig. 2**.

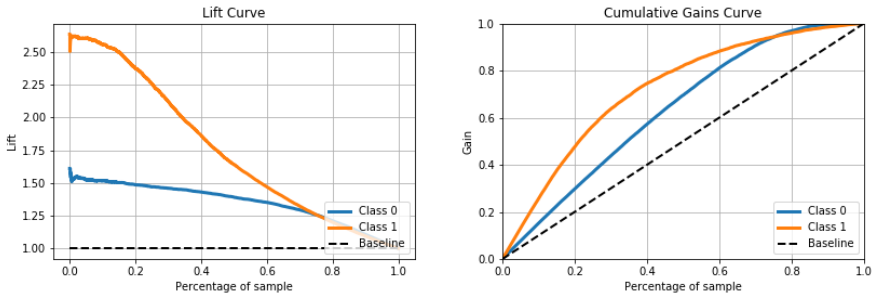

Fig. 2. (a) Lift curve of lasso logistic regression model on withdrawing and fail students, and (b) Cumulative gain curve of lasso logistic regression model on withdrawing and fail students

On the left side of **Fig. 2** one can see the lift curve. On the x-axis percentage of the sample is presented, while on the y-axis lift is presented. Dropout students are presented in orange color and denoted as Class 1, while non-dropout students are presented in blue color and denoted as Class 0. Lift is the ratio of the expected probability of dropout and the probability of dropout in the dataset. Therefore, the baseline value is 1. Examples are sorted according to the probability of dropout and one can make decisions based on it. In addition, one can use the cumulative gain curve (**Fig. 2** (b)). On the x-axis percentage of the sample is presented, and on the y-axis gain, or percentage of students of the corresponding class. Dropout students are presented in orange color and denoted as Class 1, while non-dropout students are presented in blue color and denoted as Class 0.

Having this description in mind, one can contact the top N% of the students sorted by probability of dropout. For the predictive model presented in Fig. 2 (a) it would be beneficial to contact the top 20% of students. This will yield in contacting the students that are more than twice as likely to dropout compared to the overall dropout rate. More specifically, one will contact around 50% of the students that will dropout as seen in Fig. 2 (b).

These figures aid decision-makers by presenting the results of the predictive model that is more interpretable than confusion matrix and predictive performance measures such as AUC or AUPRC. There are multiple reasons for such a statement. Predictive models can use multiple performance measures for the evaluation of the model. First, using too many performance measures can be confusing for the decision-maker, since most of them are similar by definition for experts that are not a data scientist. In addition, the simplest ones, such as accuracy, precision, and recall might be inappropriate due to the selection of the decision threshold. Decision-makers would need prior education on predictive measures and their effects. Finally, the cost of errors is not the same. The cost of contacting and giving incentive to the student that will pass the exam (called false positive) is most probably a lot less than the cost of not contacting the student that will be a dropout (called a false negative). Therefore, usage of the lift curve and cumulative gain curve will give an insight to the decision-maker how many students will be contacted in percentage terms and how likely they to be a dropout is.

Next, we need to discover why the students are prone to dropout. More specifically, we answer the question of what influence dropout. Answering this question will give insight to the decision-maker about the course design, online platform, and student characteristics that can be improved and utilized in further decision-making. This is our second research question in this experiment. For the predictive model created at the beginning of the course, it is expected that demographic attributes influence the prediction model, more specifically age and self-reported social-economic status [34]. As more interaction with the learning environment is available, quiz scores and assignment grades get more influence [40, 22]. We will present several coefficients that contributed most to the prediction model, for four time frames. More specifically, one at the beginning of the course, one after the first week of the course, another after the first month of the course, and last one after 120 days (middle of the course). These time frames are selected because they are common in the literature and it is considered that decision-makers can create a campaign and possibly influence a student in a positive direction.

Coefficients of the most important features of lasso logistic regression for the withdraw students are presented in **Fig. 3**.

At the beginning of the course, the most important features are age between 0 and 35 and between 35 and 55, and "A" level of education. These features have a negative influence on dropout. This means that students that have between 0 and 35 years of age, or between 35 and 55 are less prone to be dropouts. This finding is interesting because in MOOCs this subpopulation is more prone to be dropouts [40, 31]. As a major dropout factor one can find studied credits and disability of the students. It has been noted in the literature that regardless of the type of learning studied credits influence dropout. A number of credits that course offers are correlated with the difficulty of the course. Therefore, the difficulty of the course is one of the factors of dropout [27]. In addition, if the student takes too many online courses, resulting in many studied credits, he/she will have trouble following too many courses and most likely dropout from some of them (or even all of them). It is interesting to note that interaction with a virtual learning environment is present in coefficient even at the model for the beginning of the course. Sum of clicks on content, pages, and forum are of negative influence on dropout [15, 39]. This indicates that students do tend to interact with the learning environment (i.e. reading materials, discuss the forum, etc.) in order to be prepared for the upcoming lecture. These students highly influence to achieve satisfactory results. However, the newly used aggregation measure is of interest. More specifically, the recency of interaction with the content (i.e. videos) learning environment is introduced. Its value is negative, which can be interpreted that more recent interaction with the learning environment is negatively influencing the dropout.

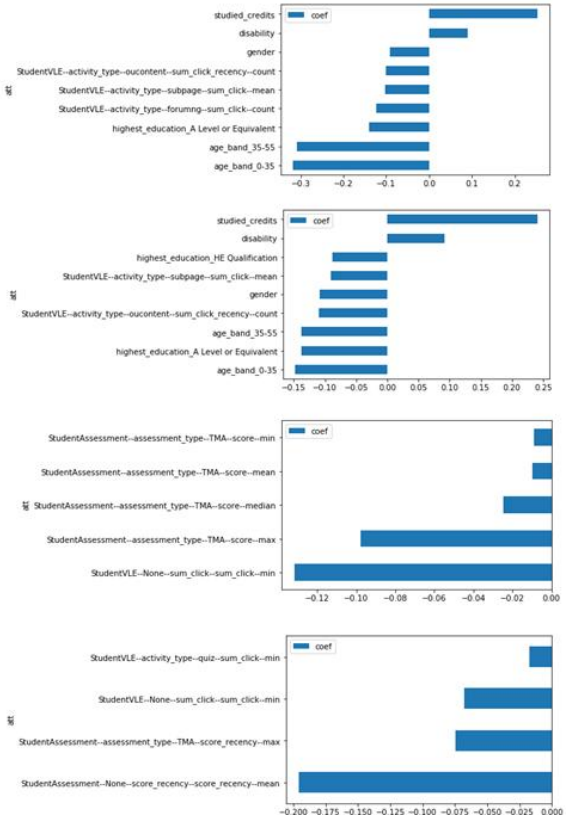

Fig. 3. Bar chart of coefficient weights for (a) beginning of the course, (b) after the first week, (c) after the first month, and (d) at the mid-term, for lasso logistic regression model for the withdraw students

After one week of lectures in the online course, the situation for the data at hand remains similar. Studied credits and disability remain the most important features for logistic regression to recognize dropouts. Also, age and activity in the virtual learning environment are most important at recognizing non-dropouts. It is worth noticing that the recency of interaction with the content (i.e. videos) learning environment remains as one of the most influential attributes in the predictive model. Although it cannot be seen in Fig. 3 variability seeking index is introduced in the predictive model at this point. More specifically, after one week of lectures students generate activities, and the variability of their activity starts making an impact on the predictions. All of the influencing attributes that utilize this aggregation function are related to the interaction with the learning environment and their values are negative. Those attributes are related to the clicking on the URLs in the supplementary materials, posting on the forum, and the overall number of clicks in the learning environment. Therefore, the positive value of variability seeking index (a positive trend in activity) leads to the passing the exam at the final of the course.

After one month of course content, the situation is drastically changed. Students have generated many activities in the learning environment and have had several quizzes and

assignments. This yielded in different patterns leading to the dropout, f.e. some of the students were discouraged by the difficulty, or some of them felt unmotivated to proceed. Attributes that were most important for the predictions in previous time periods, i.e. age, studied credits, and disability, are not important that much. They were not present in the most important attributes at all. They are replaced with scores on quizzes and assignments. Scores on tutor marked assignments are the most dominant set of features. However, the most important feature is the number of clicks in total on the virtual learning environment [38]. Recency and variability seeking index are also not present in the most important ones, but they are present in the predictive model. The recency of students' activity on the learning environment, as well as the recency of login on the learning environment and posting on the forum, are still considered as a strong negative influence on the dropout. However, variability seeking index is present only for two attributes, which are the same as for the previous model. More specifically, clicking on the URLs in the supplementary materials and posting on the forum negatively influence dropout of the students'.

After 120 days of the course (mid-course), it becomes clear what influence dropout. The students are familiar with the learning environment, the style of teaching, quizzes, and assignments. Therefore, the amount of effort that is invested in the learning environment is highly reduced, focusing only on the part of the course that results in the certificate (i.e. quizzes and assignments). Having that in mind, scores on quizzes, tutor-marked assignments, and activity on the virtual learning environment are the most important factors of dropout. However, instead of using classical aggregation function, recency is more appropriate, at least for the data at hand. As can be observed, the most influencing attribute was the recency of the obtained score. This indicates that the greater values of the recent scores are negatively influencing the dropout.

We can conclude that at the beginning of the course demographic features of the student and course description features are the most important for dropout prediction. Namely, younger students, with higher education are less prone to dropout, while the difficulty of the course and disability of the student do influence dropout. As the course goes by, student activity gains more importance. Interaction with the virtual learning environment, i.e. spending more time reading materials, posting questions and answers on the discussion forum, as well as scores on the assignments gains more importance for dropout prediction. More specifically, the higher the engagement of the student to the virtual learning environment and the higher the scores on the assignments, there is less chance that student will be a dropout. In addition, proposed aggregation functions are important for the predictive model, as the recency of activity on the learning environment and recency of scores negatively influence dropout. Variability seeking index does influence the predictive model (i.e. positive change in trend in activity on the learning environment leads to fewer dropouts), but not as strong as the recency.

The interpretation is similar, but different if dropout is defined as a student who failed the exam or withdraws from the course. Coefficients are presented in Fig. 4. At the beginning of the course, the most important feature is the "A" level of education. This feature had a negative influence on withdrawing, but in this setting (withdraw and failing the exam), it has a positive influence on dropout prediction. Besides, "A" level of education positive influence on dropout is presented in studied credits and number of previous attempts, as well as lower values of the IMD band. This effect has already been noticed by [21]. IMD band, which represent the socio-economic status of the region of the student could be that lower socio-economic status of the student influences

the performance of the student. One should be careful when interpreting this coefficient weight since lower socio-economic status surely does not cause lower performance, but that there is some confounding effect of the performance. As in previous dropout models higher the activity on the virtual learning environment, the less the probability of dropout. Additionally, proposed aggregation functions recency and variability seeking index are not present in the predictive model.

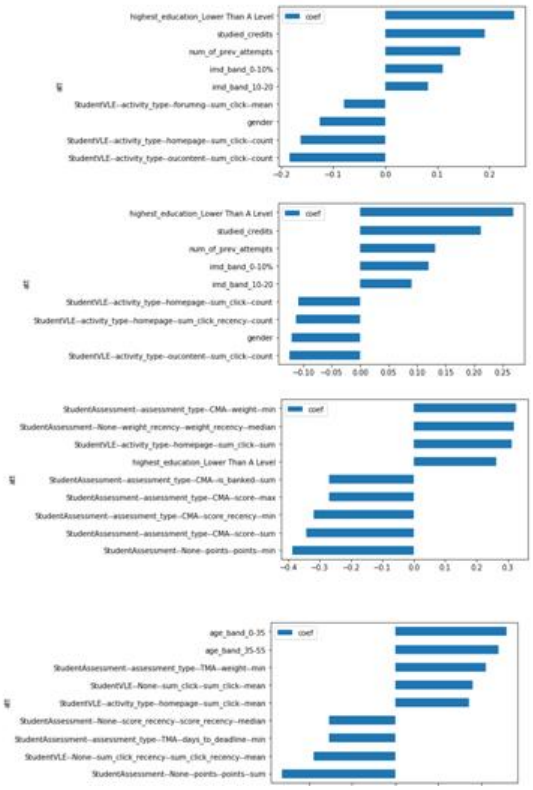

Fig. 4. Bar chart of coefficient weights for (a) beginning of the course, (b) after the first week, (c) after the first month, and (d) at the mid-term, for lasso logistic regression model for the withdraw students and fail students

The coefficients of lasso logistic regression are approximately the same for the model after the first week. More specifically, prior education studied credits, and the number of previous attempts of the course has a positive influence on the dropout predictions, while click counts on the learning materials and being female influence passing exam. However, the newly proposed recency aggregation function applied to the access to the homepage is considered a major factor that influences passing the exam. Recency had a greater influence on the prediction compared to the model predicting only withdraw students. More specifically, recency appeared in this predictive model in ten attributes, all regarding specific interaction with the learning environment (i.e. access to the forum, URLs in the learning materials, and accessing additional learning materials) with a negative value (negatively influencing dropout).

Quizzes and assignments do not influence predictions at the beginning of the course. They are introduced as important features in the model created after the first month of course. It can be observed that the higher the weight of the assignments in the first month greater the probability that students will be a dropout. Also, if a student achieved greater scores on assignments lower the probability of being a dropout. A similar finding can be seen in many dropout predictions model [7, 43]. This is because the cost of failure is low. The students tend to quit after the first several unsuccessful quizzes and assignments mostly to not feel the failure of not achieving the certificate [6]. Our newly added aggregation measures, recency, and variability seeking index have their share in the prediction. Recency, as an aggregation function, appeared as important in many attributes. First, the weight of the test. The attribute coefficient related to the recency of the weight of the test indicates that challenging tasks for the students that account for many points positively influence dropout. More specifically, obtaining a low score on the important test leads to failing the exam or withdrawing the course. Also, the variability seeking index has appeared in several interactions with the learning environment attributes. It indicates that higher usage of the learning environment leads to passing the exam.

Interestingly, after 120 days of course the age of the student returns to the most important features. It is even more surprising that this attribute has the highest positive influence on the dropout. However, it is worth to notice that recency related attributes are very important with a negative influence on the dropout. More specifically, the higher the recent score obtained it is more likely that students will pass the exam.

With the interpretation of the coefficients, we explained why do dropout occurs in general. However, for the decision support system in MOOCs one needs a detailed explanation of why a specific student did not pass the exam or what can this student do in order to pass an exam. For that purpose, we utilized counterfactual examples [30]. Those are examples that are the most similar to the real examples but having a different outcome. In the process of the dropout prediction, counterfactual examples would be students that passed the exam but are most similar to the one we would like to give incentive. By providing this kind of example to the decision-makers one could look at the attributes that can be influenced by the decision-maker in order for a student to pass the exam. Simplified example (due to the high dimensionality of the data we showed only several attributes) of counterfactual examples are presented in Table 4. Suppose we have attributes *gender*, *Forum_post* representing a total number of posts on the forum, *VLE_recency* representing a number of interactions with the virtual learning environment whose score is adjusted to valorize more recent ones, *Quiz_recency* representing average quizzes score adjusted to valorize more recent ones, and output column *dropout* that signal whether the student is a dropout. The first row of the table is the original example, while the remaining rows present counterfactual examples. *Gender*, as well as *Quiz_recency* are attributes that decision-makers cannot influence. Therefore, it will always remain the same (for all counterfactual examples). In other words, *Forum_post* and *VLE_recency* are of the point of influence to the decision-maker as those attributes can be subject to intervention. The decision-maker can request multiple counterfactual examples (in this example three) and they will slightly differ. Column *dropout* represents a signal that a student is a dropout, or probability of a dropout for counterfactual examples. In this example, in a bold letter, we have shown what strategies the decision-maker can take for communicating the student. In this simple example, it can give the incentive to interact using the forum, or interaction with

the virtual learning environment, and finally interact with both forum and virtual learning environment, but with less intensity.

Table 4. Counterfactual examples

Student	gender	Forum_post	VLE_recency	Quiz_recency	dropout
Original	1	0	0	57.2	1
CF1	1	**5**	0.57	57.2	0.12
CF2	1	0	**0.93**	57.2	0.08
CF3	1	**1**	**0.75**	57.2	0.25

Finally, once the decision-maker has the predictive model, a graphical tool for selecting the students to be contacted, interpretation of the model, and counterfactual explanation for each student that is predicted to be a dropout, one can generate a decision support system. In this paper, the most suitable solution would be the development of a module similar to a customer relationship manager (CRM). More specifically, decision-makers could create a campaign that will contact a student regularly (i.e. weekly basis) using e-mail messages and/or push notifications.

6. Conclusion

Application of data mining and machine learning in the education domain presents an interesting research area which requires a lot of technical skills (i.e. data visualization, statistics, algorithms, etc.), social skills (i.e. pedagogy, andragogy, communication skills, etc.) in order to make effective and influential decisions. In this paper, we employed lasso and ridge logistic regression for the dropout prediction of the students in the online learning environment. We asked ourselves two questions. How early can we predict dropout and can we explain why dropouts occur? Because of the vague definition of dropout, we developed two experiments where dropout was defined when student unenrolled from the course, and another when a student failed to pass an exam or unenrolled from the course.

In order to answer the first question, we created eight experiments. Namely, we created models at the beginning of the course, after the first week of the course, after 15 days, after 30 days, after 45 days, after 60 days, after 90 days, and after 120 days (mid-term). The results have shown that withdraw from the course is harder to predict. A performance measure that was selected, AUC, was 0.549 at the beginning of the course and arose to 0.681 at the mid-term. For the second definition of dropout, the performance was much greater. More specifically, AUC at the beginning of the course is 0.661 and improves up to 0.869 in the mid-term. These models can be used for informed decision making because improvement compared to uninformed decision making is from ~25% for the model at the beginning of the course, to ~71 at the mid-term.

The second research question (why do the dropout occur) is answered by analyzing the coefficients of the logistic regression model. It has been shown in both definitions that at the beginning of the course demographic features of the student such as age and education influence dropout. More specifically, younger students, with higher education are less prone to dropout. However, the difficulty of the course and disability of the student do influence dropout. Later, as students gain activity in the virtual learning

environment, the predictive model gives more importance to those attributes. More specifically, the higher the engagement of the student to the virtual learning environment and the higher the scores on the assignments, the less the probability that students will be a dropout. It is worth noticing that proposed aggregation functions recency and variability seeking index influence predictive models as they were frequently considered as one of the most important ones.

Having answers to the two proposed decision-makers can make an informed decision about contacting the troubled student. Namely, one is given the answers to the question of *who* is at trouble, and *why* is at trouble. The answer to the question of *what* to do or *how* to approach is given using counterfactual examples. In future work, we will try providing explanations using Shapley scores [4] or Lime framework [16]. Counterfactual examples do provide some notion of explanations, but Shapley scores and Lime can more human interpretable explanations.

As we believe that predictive performance is region-specific, i.e. one region has overall better results compared to the other one, we would like as a part of the future research to apply multi-task logistic regression models [46]. This type of analysis would create a predictive model for each region (one region will be one task). However, the multi-task learning framework will tend to have similar coefficients for the attributes throughout the regions. If one region is truly different in the behavior of dropouts then their coefficient for some attribute will differ (i.e. predictive strength will be much greater compared to the penalty imposed by changing the value of the coefficient). This analysis would give us the true value of the driving factors for the dropout based on the region of the student.

Another line of the research should be regarded as the problem of algorithmic fairness. More specifically, we will strive to create predictive models that are non-discriminatory or fair toward socially sensitive groups. As results suggested for the data at hand, gender seems like an attribute that discriminates passing the exam and failing to pass the exam. Since this can be an indicator of disparate impact or even disparate treatment, one should inspect why gender is making a difference in predictions. In order to make a fair predictive model and not including gender (or any other attribute that can be considered as a proxy to gender), we will try to preprocess data to be fair, adjust prediction to seem fair, or adjust the learning algorithm.

Acknowledgment. This paper is the result of the project ONR - N62909-19-1-2008 supported by the Office of Naval Research, the United States: *Aggregating computational algorithms and human decision-making preferences in multi-agent settings*.

References

1. Agarwal, S., Graepel, T., Herbrich, R., Har-Peled, S., & Roth, D. (2005). Generalization bounds for the area under the ROC curve. *Journal of Machine Learning Research*, *6*(Apr), 393-425.
2. Allione, G., & Stein, R. M. (2016). Mass attrition: An analysis of drop out from principles of microeconomics MOOC. *The Journal of Economic Education*, *47*(2), 174-186.
3. Azizah, E. N., Pujianto, U., & Nugraha, E. (2018, October). Comparative performance between C4. 5 and Naive Bayes classifiers in predicting student academic performance

in a Virtual Learning Environment. In *2018 4th International Conference on Education and Technology (ICET)* (pp. 18-22). IEEE.

4. Biecek, P. (2018). DALEX: explainers for complex predictive models in R. *The Journal of Machine Learning Research*, *19*(1), 3245-3249.
5. Bleiklie, I. (2005). Organizing higher education in a knowledge society. *Higher Education*, *49*(1-2), 31-59.
6. Chen, C., Sonnert, G., Sadler, P. M., Sasselov, D. D., Fredericks, C., & Malan, D. J. (2020). Going over the cliff: MOOC dropout behavior at chapter transition. *Distance Education*, *41*(1), 6-25.
7. Chen, Y., Chen, Q., Zhao, M., Boyer, S., Veeramachaneni, K., & Qu, H. (2016, October). DropoutSeer: Visualizing learning patterns in Massive Open Online Courses for dropout reasoning and prediction. In *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 111-120). IEEE.
8. Chengjie, Y. U. (2015). Challenges and changes of MOOC to traditional classroom teaching mode. *Canadian Social Science*, *11*(1), 135.
9. Coleman, C. A., Seaton, D. T., & Chuang, I. (2015, March). Probabilistic use cases: Discovering behavioral patterns for predicting certification. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 141-148). ACM.
10. Cui, G., Wong, M. L., & Lui, H. K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science*, *52*(4), 597-612.
11. Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233-240). ACM.
12. Delibašić, B., Radovanović, S., Jovanović, M., Obradović, Z., & Suknović, M. (2018). Ski injury predictive analytics from massive ski lift transportation data. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, *232*(3), 208-217.
13. Delibašić, B., Vukićević, M., Jovanović, M., & Suknović, M. (2012). White-Box or Black-Box Decision Tree Algorithms: Which to Use in Education?. *IEEE Transactions on Education*, *56*(3), 287-291.
14. Downes, S. (2008). Places to go: Connectivism & connective knowledge. *Innovate: Journal of Online Education*, *5*(1), 6.
15. Fei, M., & Yeung, D. Y. (2015, November). Temporal models for predicting student dropout in massive open online courses. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)* (pp. 256-263). IEEE.
16. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018, October). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 80-89). IEEE.
17. Gütl, C., Rizzardini, R. H., Chang, V., & Morales, M. (2014, September). Attrition in MOOC: Lessons learned from drop-out students. In *International Workshop on Learning Technology for Education in Cloud* (pp. 37-48). Springer, Cham.
18. Haiyang, L., Wang, Z., Benachour, P., & Tubman, P. (2018, July). A Time Series Classification Method for Behaviour-Based Dropout Prediction. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)* (pp. 191-195). IEEE.
19. Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC.
20. He, J., Bailey, J., Rubinstein, B. I., & Zhang, R. (2015, February). Identifying at-risk Students in Massive Open Online Courses. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

21. Hlioui, F., Aloui, N., & Gargouri, F. (2018, December). Understanding Learner Engagement in a Virtual Learning Environment. In *International Conference on Intelligent Systems Design and Applications* (pp. 709-719). Springer, Cham.
22. Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017, March). Ouroboros: early identification of at-risk students without models based on legacy data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 6-15). ACM.
23. Hlosta, M., Zdrahal, Z., & Zendulka, J. (2018). Are we meeting a deadline? Classification goal achievement in time in the presence of imbalanced data. *Knowledge-Based Systems*, *160*, 278-295.
24. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 112, p. 18). New York: Springer.
25. Jiang, S., Williams, A., Schenke, K., Warschauer, M., & O'Dowd, D. (2014, July). Predicting MOOC Performance with Week 1 Behavior. In *Educational Data Mining* 2014.
26. Jovanovic, M., Vukicevic, M., Milovanovic, M., & Minovic, M. (2012). Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study. *International Journal of Computational Intelligence Systems*, *5*(3), 597-610.
27. Kursun, E. (2016). Does Formal Credit Work for MOOC-Like Learning Environments?. *The International Review of Research in Open and Distributed Learning*, *17*(3).
28. Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics Dataset. *Scientific Data*, *4*, 170171.
29. Mah, D. K. (2016). Learning analytics and digital badges: Potential impact on student retention in higher education. *Technology, Knowledge and Learning*, *21*(3), 285-305.
30. Mothilal, R. K., Sharma, A., & Tan, C. (2020, January). Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 607-617).
31. Onah, D. F. (2015, June). Learners expectations and motivations using content analysis in a MOOC. In *EdMedia+ Innovate Learning* (pp. 192-201). Association for the Advancement of Computing in Education (AACE).
32. Radovanović, S., Delibašić, B., Suknović, M. (2019). How early can we predict MOOC performance? In *Proceeding of the Euro mini International Conference on Decision Support System Technology 2019* (pp. 208-214). Madeira, Portugal.
33. Ramesh, A., Goldwasser, D., Huang, B., Daume III, H., & Getoor, L. (2014, June). Learning latent engagement patterns of students in online courses. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.
34. Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning; A decision tree based approach. *Computers & Education*, *137*, 32-47.
35. Romero, C., & Ventura, S. (2013). Data Mining in Education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27.
36. Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(1), e1187.
37. Sanchez-Gordon, S., & Luján-Mora, S. (2016). How could MOOCs become accessible? The case of edX and the future of inclusive online learning. *Journal of Universal Computer Science*, *22*(1), 55-81.
38. Staubitz, T., & Meinel, C. (2018, June). Team based assignments in MOOCs: Results and Observations. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S2018)* (pp. 47-51).
39. Sunar, A. S., White, S., Abdullah, N. A., & Davis, H. C. (2016). How learners' interactions sustain engagement: a MOOC case study. *IEEE Transactions on Learning Technologies*, *10*(4), 475-487.

40. Taylor, C., Veeramachaneni, K., & O'Reilly, U. M. (2014). Likely to Stop? Predicting Stopout in Massive Open Online Courses. *arXiv preprint arXiv:1408.3382*.
41. Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273-282.
42. Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.*, *31*, 841.
43. Xie, Z. (2019). Modelling the dropout patterns of MOOC learners. *Tsinghua Science and Technology*, *25*(3), 313-324.
44. Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013, December). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop* (Vol. 11, p. 14).
45. Ye, C., & Biswas, G. (2014). Early prediction of student dropout and performance in MOOCs using higher granularity temporal information. *Journal of Learning Analytics*, *1*(3), 169-172.
46. Zhou, J., Chen, J., & Ye, J. (2011). Malsar: Multi-task learning via structural regularization. *Arizona State University*, 21.

Sandro Radovanović is a teaching assistant at the University of Belgrade, Faculty of Organizational Sciences.His main research interests are machine learning, decision support systems, decision theory, and business intelligence. So far, he published over 60 papers in journals and conference proceedings. Since 2018, he is in the Board of Assistants at the EURO Working Group on Decision Support Systems (EWG-DSS).

Boris Delibašić is Full Professor at the University of Belgrade, Faculty of Organizational Sciences (School of Management). Since 2011, he is in the Coordination board at the EURO Working Group on Decision Support Systems (EWG-DSS). His main research interests are decision support systems, machine learning algorithm design, business intelligence, and multi-attribute decision making.

Milija Suknović is Full Professor at the University of Belgrade, Faculty of Organizational Sciences (School of Management).His main research interests are decision theory, decision analysis, decision support systems, machine learning algorithm design, and business intelligence.

Received: September 20, 2020; Accepted: January 10, 2021.