Looking for a dropout predictor based on the instructional design of online courses*

Salvador Ros [0000-0001-6330-4958] and Agustín Caminero[0000-0001-9658-9646]

Spanish Distance University, UNED. Spain

Abstract. Dropout is probably one of the most critical concerns in higher education institutions. The research literature in this field is focused on asking questions about which reasons are the most frequent for dropping out, but there is a lack of research related to influence of the instructional design to find a new predictive indicator of dropout. In this work, we analyze four datasets obtained from the result of different assessment scheduled in the instructional design of the selected courses. In-depth research allows affirming that the dropout-risk group is formed by a large group of students that decided not to take one of the proposed assessment. This predictive indicator is easy to implement and allows faculty and institutions to design educational and administrative process to understand and help the students.

Keywords: dropout, instructional design, education.

1 Introduction

Dropout is probably one of the most critical concerns in higher education institutions. It is well established that this effect is especially significative in the introductory courses having the highest dropout rates[1]. Some works reported that only 67% of students complete their first-year courses [2]. This situation is common both for face to face institutions as distance or online ones.

The research literature in this field is focused on asking questions about which reasons are the most frequent for dropping out. The research community agree that the lack of time and the lack of motivation. are the most significant. However, both of these reasons were affected by many factors [2]. Some researchers are focused on building strategies for preventing abandonment (e.g. inclusion of in-game assessments or activities) [2] while others try to predict the dropout. For this purpose, the researchers look for different features to apply sophisticated algorithms to obtain a good result in the predictions. In online education and especially at Spanish Distance University, UNED exists a methodology for improving the learning process. This methodology is

based on the continuous assessment, which implies an assessment roadmap that students have to follow until the end of course.

Our research question RQ1 is related to this fact:

Is it possible from the continuous assessment roadmap to detect the different student's group exposed to dropout along the course?

At this moment, we are not interested in detecting the possible motivations of the students for deciding to drop out but to detect them to start a process to understand and help them to continue the studies. Therefore, we are looking for an indicator that allows institutions to start theses processes and understand better this phenomenon.

The rest of this paper is organized as follow. Section 2 presents some results related to dropout situations. In Section 3, we describe our -experiment to detect dropout. Section 4 presents a discussion of the results. Finally, Section 5 outlines the conclusions.

2   Related works

The drop out phenomenon is a concern in many educational institutions. Some reports prevent the high rates of drop out, especially in the introductory courses [2]. They try to understand the aspects that push students to decide to drop out or by the contrary, which factors influence the success of students. In [3], they analysed the influence of twelve factors of success in a computer science course, and they detected the three most relevant ones were: comfort level, math, and attribution to luck whereas in [4], they detected as good predictors of students´success, the student effort and comfort.

Other authors look for different strategies, focusing on preventing dropout instead of looking for different relevant factors. In [1], they concluded that using different activities related to constructivism theories as games improve the motivation and in turn, the success.

Since motivation is considered a relevant feature in the dropout, other researchers have been working in predictive models related to the grade of the motivation of the students [5]. In [6] uses Social Cognitive learning theory as a frame for a better understanding of learner motivation and analyse the time variable as a predictive variable of the motivation.

These studies are based on the selection of a set of significative features most of them specially designed for a traditional classroom and very difficult to observe in an online or distance context since they take into account information that it is impossible to obtain from the online experience. In some cases, these studies collect features obtained from logs of different systems or programs used as a complement in the subject.

Once the most relevant features are detected, they try to build a model for predicting the success or the dropout.

Finally, some researches are involved in educational strategies to avoid dropout rates. For this purpose, they need a right and early enough prediction of dropout to apply the different interventions [7].

By another hand, there is a lack of research about the instructional design´s influence for predicting or preventing the dropout. This influence is not taken into account because most of these studies are not designed for online or distance education. Finally, the irruption of new techniques based on the deep-learning paradigm is being used for the prediction of learners' engagement base on new features, (e.g. facial expression) [8].

In this paper, we present a very simple dropout predictor that uses student activity feature based on the course instructional design. This predictor allows us an early enough prediction of which students are at high risk of dropout and allows institutions and faculty to implement different educational and strategies to obtain valuable information to understand and help students.

3   Detecting dropout based on instructional design.

For answering our research question, we take as starting point the instructional design of the courses based on the methodology followed at Spanish Distance University, UNED. The instructional design of courses allows collecting different data that contain new knowledge that could be useful for improving the teaching process. Despite other consideration, if we focus in the assessment as part of the intervention in the subject, we realize that the subjects have a schedule assessment plan that allows getting information about the performance of the students in different periods. This instructional design of the courses implies a set of assessments that they have to deliver along the course called Continuous Assessment Tests, (PEC's), and a final face to face assessment, (F2F). Each assessment contributes to the final grade, but they are not compulsory. An example of this assessment design is described in **Table 1**. It describes the courses selected for our experiment. Therefore, the time variable appears as a new element to take into account the purpose of design a continuous dropout detection system.
On the other hand, we are interested in a system that is effective, feasible, useful and very easy to implement.
For this purpose, we focused on the students' attendance to the different assessments, and we define the drop out concepts used in this work. We distinguish between two kinds of dropout: Partial dropout that includes any student who didn´t take one of the assessments at the moment they had to do (PEC's). This way, a student can be in partial dropout along the course different times. The second definition is Final dropout that includes any student who didn´t take the final assessment (F2F), regardless of whether he/she took any PEC before.

3.1   Method

We have collected data from two different computer science degree subjects delivered at Spanish Distance University, UNED. We have collected data from the year 2019 and

2020. The selected courses are CS01 introductory subjects. The first one is a CS1 Database course and the second one is a CS01 Computer Networks course.

These both courses have in common the methodology of assessment. These courses have a schedule assessment plan based on a continuous assessment framework. (see **Table 1**).

Table 1. Assessment´s Contributions to final Grade

Course	PEC1	PEC2	PEC3	F2F
Database	10%	5%	15%	70%
Computer Networks	10%	10%	-	80%

We collected the academic results of the two subjects for two years at different periods in the term following the assessment's schedule.

For Database course, the first assessment was PEC1, with a weight of 10% in the final grade. The second was PEC2, with a weight of 5% in the final grade, the third assessment, PEC3, was a practical exercise with a weight of 15% in the final grade and finally, the final face-to-face exercise (F2F) with a weight of 70% in the final grade.

In the case of Computer Network course, the first assessment was PEC1 with a weight of 10% in the final grade. The second was PEC2 with a weight of 10% in the final grade. Finally, the final face-to-face exercise (F2F) with a weight of 70% in the final grade. This structure allows as calculate the partial and final dropout.

To answer the research question RQ1, we focused mainly on the fact of whether a student had taken the assessments than in the academic results collected in the different moments along the course. This selection hides, likely, other factors that could influence the motivation or engagement of students. The analysis of these factors is out of the scope of this work. We are focused on the tracking of student activity based on the instructional design to detect the set of students at risk of dropout automatically.

3.2 Results

To analyze the results, we summarized them in four tables. Each column codifies the information about how many students (rate %) followed a sequence of taking assessments along the course. (e.g. Column 001 for Computer Networks means the rate of students that didn´t take PEC1 and didn´t take PEC2 but take F2F).

Analyzing the students' rates for Computer Networks 2019 course collected in the three-assessment schedule, we can distinguish three-point of analysis. First point after PEC1. At this point, we found 43.16% of partial dropout. After PEC2, there is an increment of up 56.8%, and finally, the final dropout rate was 50.66%, (see Table 2). For the year 2020, we obtained similar results. After PEC1, the partial dropout rate is

53.56%. After PEC2 it is increased up 58.04%, and the final dropout rate was 52.77%, (see Table 3)

Table 2. Student Rates for 2019 Computer Networks subject. (PEC1,PEC2,F2F)

000	001	011	010	100	101	111	110
33.33%	5,86%	2,91%	1,06%	9,55%	8%	32,53%	6,67%

Table 3. Student Rates for 2020 Computer Networks subject. (PEC1,PEC2,F2F)

000	001	011	010	100	101	111	110
42,21%	6,86%	3,43%	1.05%	5,54%	3,43%	33,50%	3.95%

Regarding 2019 Database course, (see **Table 4**), we had a four-assessments schedule. The Partial dropout rates for the PEC´s were respectively, 28.51%, 39.25%, and 50.82%. The final dropout rate was 44.21%. The 2020 Database course data, (see **Table 5**) showed that the PEC's partial dropout rates were respectively 32.25%, 38,71%, and 62.21%. The final dropout rate was 41.01%

**Table 4.** Student rates for 2019 Database Subject. (PEC1,PEC2,PEC3,F2F)

0000	0001	0010	0011	0100	0101	0110	0111
19.00%	1.65%	3.72%	1.65%	1.65%	0%	0.42%	0.42%

1000	1001	1010	1011	1100	1101	1110	1111
6.61%	3.72%	2.48%	0.42%	4.96%	13.22%	5.37%	34.71%

Table 5. Student rates for 2020 Database Subject. (PEC1,PEC2,PEC3,F2F)

0000	0001	0010	0011	0100	0101	0110	0111
25.33%	2.80%	0.46%	0.46%	0%	0.92%	0.92%	1.38%

1000	1001	1010	1011	1100	1101	1110	1111
4.14%	3.22%	0.92%	1.38%	6.45%	19.35%	2.80%	29.47%

4   Discussion

The first trend we can observe is that the partial dropout rates increase along the course, but the final dropout rate at the end is less than the last partial dropout rate. For all the dataset used the final dropout is higher than the first partial dropout rate at PEC1, (see **Fig. 1** To establish some kind of relationship among the partial dropout rates and the final one we have to deep in the composition of these rates.
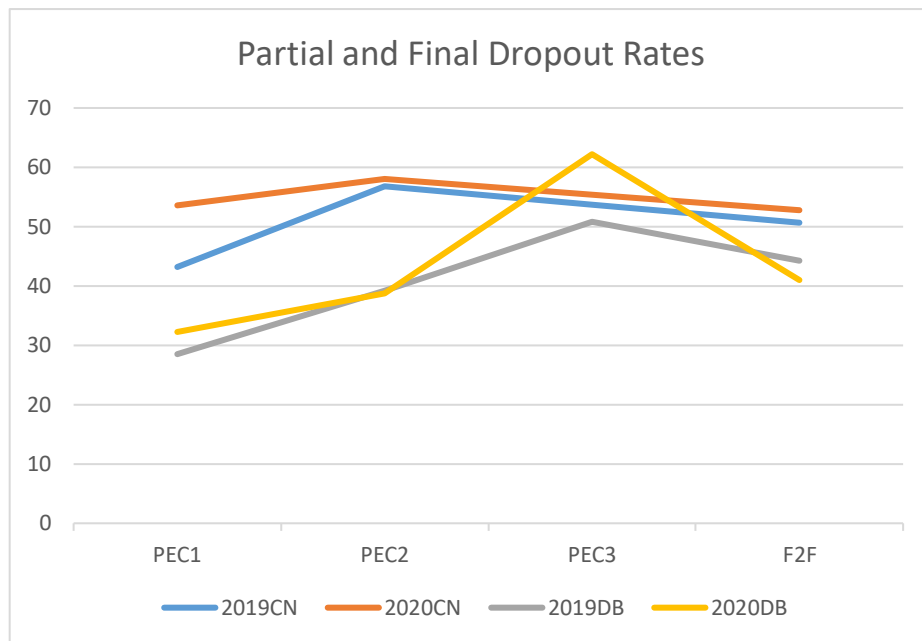
Fig. 1. Evolution of the dropout rates

Based on the 2019 Computer Networks dataset, our initial candidate dropout-risk group had a size of 43.16% of the total student. We observed that this group by the time of the PC2 was increased up 56,8%. At this point, this group was formed by 39.19% of students have taken neither PEC1 nor PEC2 and a student´s group that decided not to take the PEC2 (17,55%). Finally, a small portion of students that was initially in the dropout-risk group decided to take PEC2 and then to go out from it. In the end, the group of risk determined by the final dropout rate was formed in 65.79% by the students that had taken neither the PEC1 nor PEC2 (33,33%). We found this behaviour in the 2020 computer network dataset, as well. In this case, the final dropout-risk group was formed by a 79,98% of the students that had taken neither the PEC1 nor PEC2 (42,21%).

Finally, we would like highlighting that if we extended the dropout-risk group with the students that only have taken one PEC, the final dropout group for 2019 Computer Network course data set was formed by an 86,67% of these students and the 2020 data set by a 92,47% od these students.

After this first analysis, we wanted to check if an increase in the number of PEC´s had a significative influence on the composition of the dropout-risk group. For this purpose, we analysed the dataset for the 2019 Database course. We found that 51.39% of the final dropout-risk group is composed of students that have taken neither PEC1 nor PEC2. This rate increased up 76.63% if we add the group of students that has only take one of these two PECs. The behaviour of the 2020 Database course dataset was similar. The 62,88% of the final dropout-risk group was composed by students that have taken neither PEC1 nor PEC2, and this rate increased up 77,46% if we added the group of students that only had taken one of these two first PEC´s.

Following the analysis, and adding the third PEC, we found that the composition of the dropout-risk group for 2019 Database course increased up 87,85% if we considered the student that haven´t take one PEC. This rate increased up 93,19% in the 2020 data set for the Database course.

Therefore, the tracking of the students that didn´t take any of the proposed PEC is a good predictor of dropout group. This predictor could enable instructors to identify students at risk of dropping out and defining different teaching strategies to target the groups they belong initially formed at the beginning of the course.

5   Conclusions

In this work, we have analyses the dropout phenomenon taking into account the information obtained from the instructional design of an online course. The instructional design allows to schedule some interventions along the course. Assessments are considered a kind of intervention. Analysing, the data about how many students take the different assessments, we realized that we had a good predictor of the composition of the dropout-risk group.

This predictor gives us an initial set of students in dropout risk that allows faculty and institution to do some interventions to understand and help students. This group is

redefined each time that the students decide not to take an assessment to the extent that the results point out that the students that decide not to do one of the scheduled assessment are candidates to drop out finally.

This predictor is an effortless and efficiency one and allows an easy implementation and tracking.

6   Acknowledgements

7   Software

The dataset and the code for obtaining these results are available at https://github.com/sros-UNED/LASI2020.

References

[1]   A. Yan, M. J. Lee, and A. J. Ko, "Predicting Abandonment in Online Coding Tutorials," *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pp. 191–199, Oct. 2017, DOI: 10.1109/VLHCC.2017.8103467.

[2]   L. Malmi, "Why students drop out CS1 course," in *Proc. Second International Workshop on Computing Education Research*, 2006.

[3]   B. Wilson and S. Shrock, "Contributing to success in an introductory computer science course: A study of twelve factors," presented at the ACM Sigcse Bulletin, Mar. 2001, vol. 33, pp. 184–188, DOI: 10.1145/366413.364581.

[4]   P. R. jr Ventura, "Identifying predictors of success for an objects-first CS1," *Computer Science Education*, vol. 15, no. 3, pp. 223–243, Sep. 2005, DOI: 10.1080/08993400500224419.

[5]   L. Qu and W. L. Johnson, "Detecting the Learner's the Motivational States in An Interactive Learning Environment," in *Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*, NLD, May 2005, pp. 547–554, Accessed: May 02, 2020. [Online].

[6]   M. Cocea, "Assessment of Motivation in Online Learning Environments," in *Adaptive Hypermedia and Adaptive Web-Based Systems*, vol. 4018, V. P. Wade, H. Ashman, and B. Smyth, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 414–418.

[7]   S. Halawa, D. Greene, and P. J. Mitchell, "Dropout Prediction in MOOCs using Learner Activity Features," p. 21, 2014.

[8]   M. A. A. Dewan, F. Lin, D. Wen, M. Murshed, and Z. Uddin, "A Deep Learning Approach to Detecting Engagement of Online Learners," 2018, pp. 1895–1902, DOI: 10.1109/SmartWorld.2018.00318.