

Predicting Deadline-Driven Learners and Dropout in MOOCs: An Analysis of Learners' Behaviors

Pedro Manuel Moreno-Marcos^{1,*}, María Cantón Rello¹, Carlos Alario-Hoyos¹, Pedro J. Muñoz-Merino¹, Iria Estévez-Ayres¹ and Carlos Delgado Kloos¹

¹Department of Telematic Engineering, Universidad Carlos III de Madrid, 28911 Leganés (Madrid), Spain, ROR: 03ths8210

Abstract

The wide collection of data on digital platforms has enabled researchers to conduct many analyses around students' behaviors. One specific relevant context is the one of Massive Open Online Courses (MOOCs), where dropout is very frequent and other related behaviors such as deadline-driven strategies could be analyzed. This work aims to analyze learners' behaviors, such as deadline-driven behaviors and help seeking, and carry out predictive models to forecast how early learners will deliver their tasks and dropout. Results show that deadline-driven behaviors were frequent (about one third of learners delivered the tasks in the last 48 hours) and they had a moderate correlation with dropout. Moreover, predictive models served to predict the number of days learners delivered their tasks in advance with RMSEs between 1.1 and 1.9 days, depending on the task. As for help seeking, this was frequent in videos but not with the forum. In addition, when analyzing dropout, results showed accurate results (AUC up to 0.94) although variables about videos and formative exercises did not perform so well by themselves, without previous summative grades.

Keywords

Learning Analytics, MOOCs, Prediction, Students' behaviors, Dropout

1. Introduction

Massive Open Online Courses (MOOCs) allow the collection of data about their learners, which can be very useful to understand learners' behaviors. One typical research line is prediction in MOOCs [1], where many articles have focused on predicting dropout (or success, which is very related) due to low completion rates (even below 10% [2]). These works have used different approaches, including traditional machine learning algorithms (e.g., [3]) and deep learning (e.g., [4]). In addition, many variables have been used for those models. Among them, variables related to interactions to exercises usually achieve very powerful results (e.g., [5]), and variables related to interactions with exercises are also very common (e.g., [6]). Moreover, some other works have used variables related to forum with contrasting results about their usefulness (e.g., [7, 8]).

Some other works have tried to include more complex variables that capture learners' behaviors. Among them, Moreno-Marcos et al. [9] reported that self-regulated variables could have an impact on prediction. In addition, Liu et al. [10] focused on emotional and cognitive engagement and developed detectors that were used to predict learning achievement. Nevertheless, there are many possible learners' behaviors that may affect dropout, and the analyses of new behaviors and the development of new dependent variables based on these behaviors are also worth studying.

Among those possible behaviors, the moment when learners deliver their task may also have an impact. In this direction, some works have analyzed procrastination, and e.g., Yao et al. reported

LASI Spain'25: Learning Analytics Summer Institute, May 26–27, 2025, Vitoria, Spain

*Corresponding author.

✉ pemoreno@it.uc3m.es (P. M. Moreno-Marcos); 100406706@alumnos.uc3m.es (M. Cantón Rello); calario@it.uc3m.es (C. Alario-Hoyos); pedmume@it.uc3m.es (P. J. Muñoz-Merino); ayres@it.uc3m.es (I. Estévez-Ayres); cdk@it.uc3m.es (C. Delgado Kloos)

🌐 <https://www.it.uc3m.es/pemoreno/> (P. M. Moreno-Marcos); <https://www.it.uc3m.es/calario/> (C. Alario-Hoyos); <http://www.it.uc3m.es/~pedmume/> (P. J. Muñoz-Merino); <https://www.it.uc3m.es/iria/> (I. Estévez-Ayres); <http://www.it.uc3m.es/cdk/> (C. Delgado Kloos)

>ID 0000-0003-0835-1414 (P. M. Moreno-Marcos); 0000-0002-3082-0814 (C. Alario-Hoyos); 0000-0002-2552-4674 (P. J. Muñoz-Merino); 0000-0002-1047-5398 (I. Estévez-Ayres); 0000-0003-4093-3705 (C. Delgado Kloos)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

connections between past and future activities [11] and highlighted that this effect was prevalent in MOOCs [11], probably influenced by their relationship with self-regulated learning. In an undergraduate context, Nicholls [12] also reported a significant negative relationship between procrastination and grades. In a more general way, it is possible to refer to deadline-driven behaviors as learners do not necessarily procrastinate when they deliver their task late. In this line, Sheshadri et al. [13] reported that learners are deadline-driven even in blended courses and Nabizadeh et al. [14] carried out an analysis in a gamified course and also reported that most of the students were deadline-driven and became more active at the end of the semester. However, more work is needed to specifically analyze deadline-driven behaviors in MOOCs.

Another interesting behavior is help seeking. This behavior has been widely analyzed in Intelligent Tutoring Systems [15] where the systems usually offer hints, but it is interesting to analyze how learners do their assessments in the MOOC. For example, research has shown that many learners use the strategy of opening the assessments first to then look for the questions in the videos [16]. Similarly, it would be relevant to analyze whether or not learners look for information in the forums while taking the exams and the relationship of these behaviors with dropout.

Considering this context, the objective of this work is to contribute with an analysis of deadline-driven behaviors and help seeking, and how to predict both deadline-driven behaviors and dropout using multiple features, including help seeking.

The remainder of the paper is as follows. Section 2 presents the methodology of the paper, including details about the MOOC design, the independent and dependent variables included in this study, and the analytical methods and metrics. Section 4 details the results of this paper, including the results derived from a descriptive and a predictive analysis. Finally, Section 4 provides the conclusions of this paper and discusses the main limitations and possible future research directions.

2. Methodology

The study was carried out using data from a MOOC with five weekly modules on programming hosted by Universidad Carlos III de Madrid. This course was instructor paced and consisted of five weekly multiple-choice tests (15% of the grade each) and two peer-review programming assignments (25% of the grade), which were done in weeks 3 and 5. The passing rate of the MOOC was 60% and it was not necessary to complete all the assignments to pass the course. The total duration of the course was 8 weeks, considering that each module was opened for two weeks (except the first one, which was opened for three weeks). Regarding the possible activities, the course included multiple videos and formative questions to prepare for the assessments, and a forum to ask questions. Considering the interactions with these materials, several variables were computed for this analysis. The full list of them is presented in Table 1.

For the predictive models, the variables to predict are the number of days in advance learners submit the tasks (deadline-driven behavior, *diff_ex*) and prevail (used in this study to represent dropout), and the rest serve as predictor variables. Models are evaluated using Decision Trees (DT), Random Forest (RF), k-Nearest Neighbors (kNN), and Support Vector Machines (SVM). For the deadline-driven behavior analysis, Root Mean Square Error (RMSE) is used, and for the dropout one, Area Under the Curve (AUC) is computed to evaluate the results. For the first case, RMSE is used as the dependent variable is continuous and this metric might be better than others for this type of research, according to the literature [17]. Moreover, AUC is used for dropout as this metric is generally appropriate for student behavior classification problems [17].

3. Results

This section is divided into two subsections to present a descriptive analysis of deadline-driven behaviors and a predictive analysis of deadline-driven behaviors and dropout.

Table 1

List of variables involved in the analyses

Variable	Description
ex_x	It indicates the grade of each exam, where x is the week where the exam belongs (e.g., exam_1, exam_2, etc.)
n_c_e_x	It indicates the number of completed exams considering the specific attributes such as number of questions completed, and the specific week x.
form_x	It indicates the grade of formative questions in week x
n_c_f_ex	It indicates the number of chapters that students complete the individual formative exercises in week x
watched_x	It quantifies the amount of videos a student has engaged with in week x. This includes all videos with interactions, even if they are not finished
uninterrupted_x	It measures the quantity of videos that the student has watched from start to finish, without interruption, in week x.
style_ex	Binary variable indicating if learners watch the videos and take the exams afterwards or they go to the exams first in week x
hsv_x	It is used to represent help-seeking and it indicates whether or not learners have engaged with videos during the exam (between the first and last exam interaction) in week x
hsf_x	It is used to represent help-seeking and it indicates whether or not learners have used the forum to seek help while taking the exams
diff_ex	It represents the deadline-driven behavior and it indicates the number of days between exam completion and due date in week x.
prevail	Binary variable indicating whether or not learners have completed all the exams

3.1. Descriptive analysis

The first analysis consisted of determining how early learners usually delivered their tasks. To do that, a histogram with the distribution of the number of days in advance that learners delivered the MOOC assignments was calculated. This histogram is shown in Figure 1. From this figure, it can be inferred that roughly 22.8% (4487 students) delivered their tasks within the last 24 hours, and 10.23% between the last 24 and 48 hours.

While this is not necessarily negative since many learners may organize well to do it in the last days, it is also interesting to delve into how delivering early may affect dropout. In order to do that, the number of learners who drop out the course the following week when they deliver the task late is analyzed. Table 2 indicates the number of learners who deliver each task within the last 24h and between the last 24-48h, and the number of them who drop out at next exam.

In the first exam, results show that not many learners (considering the total) delivered the tasks in the last moment, probably due to their enthusiasm to start the course. However, as time passes, the percentage of students showing deadline-driven behaviors increased, which could highlight lack of time, happiness with the course or an increased/reduced task difficulty, which is a situation that has been previously observed in the literature [18]. It is interesting that the number of learners who showed deadline-driven behaviors was higher in Week 3. A possible reason is that the first peer-review activity was due that week and learners had to organize well to cover both the exam and the peer-review lab activity. Moreover, it is interesting that almost 50% of the learners who submitted their tasks in the last 24h in the first two tests dropped out the course by the next examination, and this behavior has a moderate correlation with prevail (0.41, i.e., learners who submitted their tasks with more days in advance were less likely to drop out). In addition, 40% of the learners who submitted between 24-48h in advance also dropped out. These figures decreased in the next exams and only 36% of the learners who delivered in the last 24h in Exam 4 dropped out the course in the last exam.

Finally, regarding help seeking, Figure 2 shows the number of interactions showing help-seeking behaviors when learners used videos or the forum while taking the exams. Results show that the level of forum help-seeking is almost negligible, although it was more frequent for the case of videos. This shows that videos could be a valuable source of information that learners might use to clarify their

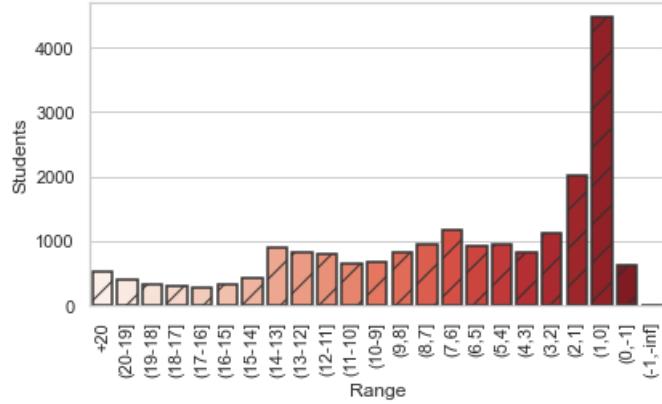


Figure 1: Distribution of the number of days students deliver the exams before the deadlines

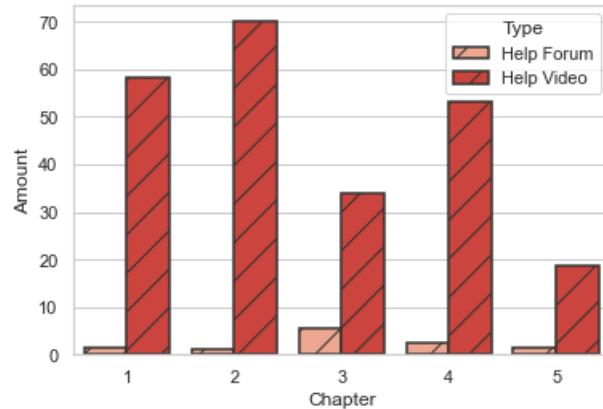


Figure 2: Comparison of Video vs Forum Help

Exam	Deliver <24h	Deliver 24-48h	Total Students	<24h Drop Out	24-48h Drop Out
1	1388	565	7887	668	224
2	1259	430	4596	598	174
3	1133	355	3229	438	107
4	463	542	2532	168	136
5	244	127	1428	-	-

Table 2

Analysis of deadline-driver behavior and their relationship with drop out for each exam

concepts while taking the MOOC assessments. Nevertheless, a limitation of this figure is that learners might also use other sources (e.g., materials outside the platform) to seek help during the task, and this is not tracked. Further analysis should be done around this.

3.2. Predictive analysis

As for the predictive models, weekly models to predict the average days of anticipation in submissions (deadline-driven behavior) and drop out were developed using the algorithms mentioned in Section 3.

For the prediction of deadline-driver behaviors, four situations were analyzed and five models were developed for each week and algorithm. These situations (S) are:

- S1. The average number of days of anticipation in submissions (considering all five weeks) is predicted using weekly cumulative data (i.e., variables in week 3 include interactions in weeks 1, 2 and 3 to predict the dependent variable referred to the average of weeks 1-5).

Week	DT				RF				kNN				SVM			
	S1	S2	S3	S4												
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	1.44	2.38	1.51	2.73	1.33	2.56	1.59	2.32	1.36	1.93	1.67	1.87	1.43	2.31	1.25	2.01
3	1.42	1.66	1.79	1.47	1.34	1.88	1.46	1.49	1.34	1.50	1.68	1.50	1.43	1.54	1.35	1.54
4	1.34	1.48	1.75	1.34	1.33	1.74	1.46	1.33	1.41	1.35	1.78	1.35	1.51	1.53	1.45	1.53
5	1.39	1.31	1.65	1.28	1.34	1.67	1.16	1.12	1.41	1.16	1.71	1.16	1.56	1.16	1.55	1.15

Table 3

Prediction of deadline-driven behaviors (values for each algorithm and situation, S, provided in RMSE)

- S2. The weekly number of days of anticipation in the submissions is predicted using weekly cumulative data (i.e., variables in week 3 include interactions of weeks 1, 2 and 3 to predict the dependent variable referred to week 3).
- S3. The average number of days of anticipation in submissions (considering all five weeks) is predicted using weekly independent data (i.e., variables in week 3 include interactions in week 3 to predict the dependent variable referred to the average of weeks 1-5).
- S4. The weekly number of days of anticipation in the submissions is predicted using weekly independent data (i.e., variables in week 3 include interactions in week 3 to predict the dependent variable referred to week 3).

Considering these situations, results of the predictive models are presented in Table 3. Results show that it is possible to detect the number of days of anticipation in tasks with an RMSE below 2 days in most of the cases. Considering situation 1, it is possible to detect the overall number of days of anticipation with an RMSE of 1.3 in all the weeks (except the first one, which cannot be predicted as there is no previous data) and the best results are obtained using RF. When comparing situations 1 and 3, which predict the same dependent variables, results are generally better when using cumulative data, although the best overall value in week 5 is with RF and situation 3. However, when predicting specific values in specific exams, most recent data provided better results in most of the cases. In this case, the reported RMSEs are between 1.1 and 1.9 days depending on the task, and while the best result in week 2 is achieved with cumulative data (situation 2) and kNN, values obtained in situation 4 stand out in the rest of the cases. Regarding the algorithm, RF stands out in weeks 4 and 5, while DT achieves the best results in week 3. Thus, RF stands out in most of the cases considering the different cases.

Regarding the dropout analysis, several models were also defined depending on several situations (S) to analyze the predictive power using multiple sets of variables. Particularly, the situations defined for this case (all of them using cumulative mode) are:

- S1. This model includes all the variables in Table 1 (*prevail* is used as dependent variable and the rest as independent variables).
- S2. This model includes variables related to interactions with videos to predict dropout (*prevail* variable). The independent variables are *uninterrupted_x*, *watched_x*, and *hsv_x*.
- S3. This model includes variables related to interactions with formative activities (*form_x* and *n_c_f_e*) to predict dropout ((*prevail*).
- S4. This model includes variables related to interactions with videos and exam grades to predict dropout (*prevail* variable). The independent variables are *uninterrupted_x*, *watched_x*, *hsv_x*, and *ex_x*.
- S5. This model includes variables related to interactions with formative activities and exam grades (*form_x*, *n_c_f_e* and *ex_x*) to predict dropout ((*prevail*).

Considering these situations, several models have been implemented and results are presented in Table 3. From this table, it can be seen that it is possible to obtain an AUC up to 0.94 using all variables (and above 0.8 from week 3). However, variables on video interactions (including help seeking using

Week	DT					RF					kNN					SVM				
	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
1	0.69	0.50	0.50	0.58	0.54	0.61	0.52	0.50	0.59	0.54	0.59	0.56	0.50	0.59	0.54	0.50	0.50	0.50	0.50	0.50
2	0.75	0.60	0.53	0.71	0.74	0.73	0.60	0.53	0.73	0.73	0.66	0.62	0.67	0.72	0.69	0.60	0.50	0.50	0.57	0.73
3	0.81	0.61	0.70	0.80	0.81	0.88	0.64	0.69	0.84	0.85	0.74	0.63	0.70	0.84	0.86	0.69	0.60	0.50	0.63	0.87
4	0.88	0.71	0.72	0.88	0.88	0.92	0.72	0.73	0.91	0.90	0.82	0.71	0.61	0.89	0.92	0.80	0.71	0.81	0.73	0.93
5	0.94	0.70	0.75	0.74	0.92	0.94	0.72	0.76	0.92	0.93	0.86	0.83	0.79	0.92	0.94	0.86	0.72	0.84	0.74	0.94

Table 4

Dropout prediction (values for each algorithm and situation, S, provided in AUC)

videos) and formative assessments were not enough (AUCs up to 0.83 and 0.84, respectively), although they increased up to 0.92 and 0.94 when adding previous summative grades, respectively. This result indicates that previous performance is a key factor for predictive models, as reported in other works (e.g., [9, 5]. Regarding variables related to videos and formative activities, results show that while they may have an effect on dropout, their predictive power is worse. In fact, variables related to interactions with videos only achieve an AUC higher than 0.8 in week 5 with kNN, which is not enough to anticipate results and have an impact on learners. Similarly, the AUC when using variables related to interactions with formative activities is above 0.8 only from week 4 (80% of the course duration) with SVM, which limits the possibility of obtaining accurate early predictions. With regard to the algorithms, more variability was found depending on the conditions although the best overall result was obtained with RF.

4. Conclusions

This work has analyzed the effect of deadline-driven behaviors and help seeking and the predictive models of deadline-driven behaviors and dropout. Deadline-driven behavior was found to be a common behavior in MOOCs as about one third of the learners delivered their tasks in the last two days. In addition, it was found that these behavior have an impact on dropout and e.g., almost 50% of the learners who submit their tasks in the last 24h in the first two test drop out the course by the next exam. Furthermore, predictive models were developed to forecast the number of days in advance learners submitted their tasks. Results showed that it is possible to achieve accurate predictions with an RMSE between 1.1 and 1.9 days depending on the task, and an overall RMSE of around 1.3 days when predicting the average value of all tasks. As for the the case of help seeking, help was mainly sought through videos, and there were not many identified cases of help seeking using the course forum. With regard to drop out prediction, accurate models were obtained, particularly when using previous summative grades. Particularly, an AUC up to 0.94 was achieved using all variables and above 0.8 from week 3. Nevertheless, variables related to interactions with videos and formative exercises provided accurate results only at the end of the course, which limit their impact when they are not combined with other variables related to student performance.

Despite the aforementioned findings, there are some limitations that are worth mentioning. First, results have been obtained using a single course, which limits the generalization of the findings. In addition, deadline-driven behaviors might also be impacted by the number of days available for the task, and more contexts are needed to address that. Furthermore, help seeking variables are limited as they only focus on the activity within the platform and learners may seek help from other sources. In addition, other possible models for the detection of help seeking might be used. Moreover, the definition of dropout has been established considering the completion of all exams, but other definitions could have been used and that could affect the results.

As future work, it would be relevant to analyze more contexts to discover how results may generalize to other scenarios. Moreover, further analyses could be done to analyze the reasons behind submitting the tasks near the deadlines and/or dropout. In addition, it would be relevant to track gather more data

about how learners take the exam and what additional sources they might use (including e.g., the use of Generative Artificial Intelligence). Finally, it would be relevant to carry out pilots in live courses to evaluate the applicability of these findings and how instructors can benefit of them to make an impact on their learners.

Acknowledgments

This work was supported by Universidad Carlos III de Madrid (UC3M) through the Grants for the Research Activity of Young Doctors of the UC3M's Own Research and Transfer Program (ASESOR-IA project). Moreover, it was supported by FEDER / Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación through the grant PID2023-146692OB-C31 (GENIE Learn project) funded by MICIU/AEI/10.13039/501100011033 and by ERDF/UE, by the UNESCO Chair of "Scalable Digital Education for All" at UC3M and by the grant RED2022-134284-T funded by MICI-U/AEI/10.13039/501100011033.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, C. Delgado Kloos, Prediction in MOOCs: A review and future research directions, *IEEE transactions on Learning Technologies* 12 (2018) 384–401.
- [2] S. Yin, Q. Shang, H. Wang, B. Che, The analysis and early warning of student loss in MOOC course, in: *Proceedings of the ACM Turing Celebration Conference-China*, 2019, pp. 1–6.
- [3] P. M. Moreno-Marcos, T.-C. Pong, P. J. Munoz-Merino, C. Delgado Kloos, Analysis of the factors influencing learners' performance prediction with learning analytics, *IEEE Access* 8 (2020) 5264–5282.
- [4] K. Talebi, Z. Torabi, N. Daneshpour, Ensemble models based on CNN and LSTM for dropout prediction in MOOC, *Expert Systems with Applications* 235 (2024) 121187.
- [5] Z. Ren, H. Rangwala, A. Johri, Predicting Performance on MOOC Assessments Using Multi-Regression Models, in: *Proc. 9th International Conference on Educational Data Mining*, Raleigh, NC, USA, 2016, pp. 484–489.
- [6] H. Liu, X. Chen, F. Zhao, Learning behavior feature fused deep learning network model for MOOC dropout prediction, *Education and Information Technologies* 29 (2024) 3257–3278.
- [7] R. Cobos, V. M. Palla, edX-MAS: Model analyzer system, in: *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*, 2017, pp. 1–7.
- [8] M. Klüsener, A. Fortenbacher, Predicting students' success based on forum activities in MOOCs, in: *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, volume 2, IEEE, 2015, pp. 925–928.
- [9] P. M. Moreno-Marcos, P. J. Muñoz-Merino, J. Maldonado-Mahauad, M. Pérez-Sanagustín, C. Alario-Hoyos, C. Delgado Kloos, Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs, *Computers & education* 145 (2020) 103728.
- [10] S. Liu, S. Liu, Z. Liu, X. Peng, Z. Yang, Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement, *Computers & Education* 181 (2022) 104461.
- [11] M. Yao, S. Sahebi, R. Feyzi Behnagh, Analyzing student procrastination in MOOCs: a multivariate Hawkes approach, in: *Proceedings of the 13th conference on educational data mining (EDM2020)*, 2020.
- [12] N. Nicholls, Procrastination and grades: Can students be nudged towards better outcomes?, *International Review of Economics Education* 42 (2023) 100256.

- [13] A. Sheshadri, N. Gitinabard, C. F. Lynch, T. Barnes, S. Heckman, Predicting student performance based on online study habits: A study of blended courses, in: Proc. 11th International Conference on Educational Data Mining, Buffalo, NY, USA, 2018, pp. 1–7.
- [14] A. H. Nabizadeh, J. Jorge, S. Gama, D. Gonçalves, How do students behave in a gamified course?—A ten-year study, *IEEE Access* 9 (2021) 81008–81031.
- [15] V. Aleven, I. Roll, B. M. McLaren, K. R. Koedinger, Help helps, but only so much: Research on help seeking with intelligent tutoring systems, *International Journal of Artificial Intelligence in Education* 26 (2016) 205–223.
- [16] J. Maldonado-Mahauad, M. Pérez-Sanagustín, P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, C. Delgado-Kloos, Predicting learners' success in a self-paced MOOC through sequence patterns of self-regulated learning, in: Lifelong Technology-Enhanced Learning: 13th European Conference on Technology Enhanced Learning, EC-TEL 2018, Leeds, UK, September 3-5, 2018, Proceedings 13, Springer, 2018, pp. 355–369.
- [17] R. Pelánek, Metrics for Evaluation of Student Models., *Journal of Educational Data Mining* 7 (2015) 1–19.
- [18] S. Shah, A. Mumtaz, A. Chughtai, Subjective Happiness and Academic Procrastination Among Medical Students: The Dilemma of Unhappy and Lazy Pupils., *PRAS Open* 1 (2017) 1–17.