

Article

Predicting Dropout in Programming MOOCs through Demographic Insights

Jakub Swacha *  and Karolina Muszyńska 

Institute of Management, University of Szczecin, 71-454 Szczecin, Poland; karolina.muszynska@usz.edu.pl

* Correspondence: jakub.swacha@usz.edu.pl

Abstract: Massive Open Online Courses (MOOCs) have gained widespread popularity for their potential to offer education to an unlimited global audience. However, they also face a critical challenge in the form of high dropout rates. This paper addresses the need to identify students at risk of dropping out early in MOOCs, enabling course organizers to provide targeted support or adapt the course content to meet students' expectations. In this context, zero-time dropout predictors, which utilize demographic data before the course commences, hold significant potential. Despite a lack of consensus in the existing literature regarding the efficacy of demographic data in dropout prediction, this study delves into this issue to contribute new insights to the ongoing discourse. Through an extensive review of prior research and a detailed analysis of data acquired from two programming MOOCs, we aim to shed light on the relationship between students' demographic characteristics and their likelihood of early dropout from MOOCs, using logistic regression. This research extends the current understanding of the impact of demographic features on student retention. The results indicate that age, education level, student status, nationality, and disability can be used as predictors of dropout rate, though not in every course. The findings presented here are expected to affect the development of more effective strategies for reducing MOOC dropout rates, ultimately enhancing the educational experience for online learners.



Citation: Swacha, J.; Muszyńska, K. Predicting Dropout in Programming MOOCs through Demographic Insights. *Electronics* **2023**, *12*, 4674. <https://doi.org/10.3390/electronics12224674>

Academic Editors: Miguel Morales, Roberto Barchino, José Amelio Medina-Merodio and Antonio Moreira Teixeira

Received: 16 October 2023
Revised: 10 November 2023
Accepted: 15 November 2023
Published: 16 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: MOOC; dropout; prediction; demographic features

1. Introduction

1.1. Motivation

A Massive Open Online Course (MOOC) is “an online course open to anyone without restrictions, usually structured around a set of learning goals in an area of study, which often runs over a specific period of time on an online platform which allows interactive possibilities that facilitate the creation of a learning community” [1] (p. 2). Thanks to its advantages, such as theoretically unlimited participation and open access via the Internet for anyone in the world [2], MOOCs attracted millions of online users even prior to the COVID-19 pandemic [3], which further increased their popularity tremendously [4].

While MOOCs' ability to improve academic achievements has been confirmed by numerous research reports (see the meta-analysis performed in [5]), so has its main weakness: poor student retention, which manifests as very high dropout rates. While these rates vary from course to course, typically, numbers around 95% are reported (see [3] and works cited therein), which means every 19 out of 20 students who started a course do not finish it.

Although there are various reasons for which students may leave a course (see Section 2.2 in [6] for a discussion of such reasons), some of them impossible to cope with in any way (e.g., becoming sick or heavily burdened with family responsibilities [7]), many could be addressed by an intervention from respective course organizers (see, e.g., [8]) and/or better adaptation of the course contents to students' expectations (see, e.g., [7]).

In this context, it is highly desirable to be able to identify particular students who are at high risk of dropout so that the course organizers can focus more on supporting them,

or at least identify groups of students who are at high risk of dropout so that the course can be provided in an adaptable form, with such groups of students receiving a version that differs in scope or form to the baseline version. Dropout predictors for time period $t_{n...n+1}$ can be based on the student's behavior in time period $t_{n-m...n-1}$, and any variable whose value is known in the time period $t_{0...m}$ can be considered as an early dropout predictor. Knowing that students may drop out of the course at its very beginning (see, e.g., [8,9]), there is a special value associated with variables whose value is known before the course begins, i.e., at the time point t_0 . We call these zero-time dropout predictors. Although they cannot be based on the student's prior behavior in a particular course, as the course has not started yet, there are other data sources available.

In this paper, we focus on a particular type of zero-time dropout predictor that is of demographic character. As we shall reveal in the subsequent subsections, there is no agreement in prior literature on whether students' demographic data are of any use for dropout prediction. In this paper, we strive to present new findings which enrich the existing discussion on this topic.

1.2. Problem Setting

There are different ways in which dropout can be defined [10]. In this paper, we understand this term as not obtaining a course completion certificate during its entire period of availability (i.e., from the time of course opening till its closing). This means the dropout rate includes, among others, students who passed through all units of the course but failed to pass the final test, or users who merely browsed through the course content looking for specific information they desired without actual intention to attend the course. Nonetheless, most of dropouts are usually attributed to students who joined the course with an attempt to complete it, but for some reason, did not do it, so it is generally assumed that high dropout rates in MOOCs are something undesirable and reasons for that phenomenon need to be examined and appropriate interventions undertaken to minimize dropout. Dropout prediction is a necessary means to direct such interventions to students who are at the highest risk of dropping out.

There are many studies devoted to the topic of dropout prediction in MOOCs. They apply different approaches and methods in an attempt to attain the highest possible prediction accuracy and usually benchmark their results against previously developed models [11–16]. Across the literature, different aspects influencing dropout are studied, and they can be grouped into three main categories: participant factors (including academic background, experiences, skills, and psychological attributes), course factors (including course design, institutional support, and interactions) and environmental factors (including work commitments and a supportive study environment) [17]. One of the most commonly indicated influencing factors is clickstream, which depicts participants' activity (interactions, behavior) in the online course [13,14,18–20]. Another important aspect influencing the dropout rate is the features of the course (e.g., quality and content [17,21], difficulty and length [6,11]). Less frequently studied aspects include gamification [22], the relation of registration date to the course start date [23], course start date, course length or assessment type [24], and social factors (posts on the forum) [25]. In some cases, the dropout rate was predicted based on the responses to a survey taken by course participants which, among others, indicated their interests, previous knowledge, and motivations for studying the MOOC [26]. Several studies take into account both behavioral and demographic features to increase the accuracy of online courses dropouts prediction [15,27–34]. There are also studies showing the impact of user information and course attributes on dropout prediction [16,35–38]. Throughout the analysis of the studies, behavioral (log data) followed by demographic data were the most widely used in both dropout prediction and performance prediction. Learning behavior was also the most widely utilized set of features for predicting learner dropout; demographic and assessment features were less frequently used [39].

1.3. Approach

In this paper, we put the focus on demographic characteristics of students, which are a convenient type of zero-time dropout predictor for several reasons. They are usually easily available to acquire (as course registration forms usually include some demographic questions), do not require access to external data (in contrast to, e.g., data on a student's prior behavior in other courses), and do not require students to answer any dedicated survey forms.

Such an approach, making use of demographic characteristics of MOOC participants, such as age, gender, nationality, or employment status, has been tried in many prior studies; however, their findings are not conclusive, with large differences among them regarding the reported relationship between demographics and dropout rates [17]. Some studies (e.g., [3,27,29,40,41]) prove that demographic features impact the dropout rate. On the other hand, other research (e.g., [42–44]) claims that they have no influence. This discrepancy is the primary reason we consider this research topic still worthy of our attention.

Table 1 lists identified literature sources that took into account demographic predictors influencing the dropout/completion rate in MOOCs and online courses.

Table 1. Studies including demographic predictors influencing the dropout/completion rate in MOOCs and online courses.

Source	Sample Size	Gender	Birthdate/Age	Domicile/Nationality	Employment Status	Education Level	Other Predictors
[16]	269	X	X				E1, E2, E3, E5, H2, S2, S3
[34]	597,692 *	X	X			X	E3, U1, U3, U5, U8
[36]	379	X	X		X	X	C1, C2, U4
[45]	2338	X	X	X	X	X	D9
[37]	67,333	X	X	X		X	A1, A4, H1, U5, U6, U7
[46]	79	X	X			X	D10
[30]	32,593 **	X	X	X		X	B9, D1, D11, E4, U1, U2, U7
[27]	Unspecified	X	X	X	X	X	D2–D8, H1, S1, S2
[3]	668,017 ***	X	X			X	C1, U1, U5, U9, U11
[47]	624	X	X	X	X	X	A3, A4, D6, D9, H3
[28]	32,593 **	X	X	X		X	D1, D11, E4, U1, U2, U7
[38]	1069	X	X		X	X	A4, H1, H4, S4, S5
[48]	1038	X	X			X	A3, A4, H1, S4, U11
[29]	14,791	X	X			X	D10, S2, U7, U9, U11
[49]	154,763	X	X			X	D9, U1, U7, U9, U14

* HarvardX, ** OULAD, *** XuetangX.

Codes used in Table 1 in the “Other predictors” column:

E—Educational background (E1—ACT Comp Score, E2—high school GPA, E3—current college GPA, E4—studied credits, E5—credits completed);

U—course usage (U1—no. of clickstream events, U2—no. of visits, U3—viewed/explored features, U4—interaction with the instructor, U5—videos watched, U6—assessments attempted, U7—grades achieved/assessment scores, U8—no. of chapters read, U9—no. of forum postings, U10—no. of previous attempts, U11—completed activities, U12—learning achievements, U13—no. of in-course activities, U14—time devoted);

D—Other demographics (D1—disability, D2—mother tongue, D3—current job role/occupation, D4—years of experience in the role/occupation, D5—average amount of daily working hours, D6—level of English language skills, D7—current digital proficiency, D8—number of underage children, D9—online experience, D10—marital status, D11—Index of Multiple Deprivation (IMD));

H—Student online learning history (H1—prior MOOC attendance/experience, H2—previous drops, H3—participation in online groups, H4—use of chat tools);

S—Student context (S1—available study hours per week, S2—financial aid status, S3—degree seeking status, S4—programming experience, S5—Python experience);

A—Attitude towards course (A1—intended hours per week to spend on the course, A2—satisfaction with the course, A3—motivation for taking MOOC, A4—intention of completing the course);

C—Course-specific data (C1—MOOC course content, C2—MOOC platform).

1.4. Contributions

Our contribution is threefold. First, we extend the existing collection of reports on the suitability (or not) of predicting dropout with the help of students' demographic data. Second, we identify those demographic variables that were found to be most promising in this role, according to our research results. Third, we compare results based on two similar programming MOOCs, finding outstanding differences between them, which allows conclusions to be drawn at a higher level of abstraction.

The paper is structured as follows. In the following section, we present the two programming MOOCs (one on Python, the other on JavaScript) from which the dropout data were acquired, the demographic data of their participants, and the applied research procedure. Next, we report the results of zero-time prediction of MOOC dropouts for Python and JavaScript, respectively. In the final section, the most important results are discussed.

2. Materials and Methods

2.1. Data Sources

The “Introduction to programming in Python 3” MOOC (later referred to as the Python course) was designed as an open educational resource with contents suitable for university students, as well as high school students and self-learning programming enthusiasts. The course was developed in two language versions—Polish and English—to make sure it was also available for English-speaking students (e.g., international exchange students studying in Poland). The course structure comprises three levels: modules (each ending with a module summative test), lessons, and units. Moreover, it includes an introduction with an opening test to assess the level of knowledge of Python programming at the beginning of the course and a conclusion with a closing test to assess the level of knowledge at the end of the course. No elements enforcing the pace of learning were used—the only limitation was the general time frame of the course (its beginning and end date). The list of modules constituting the Python course is presented in the second column of Table 2.

Table 2. Python and JavaScript course modules.

Lesson No.	Python Course Modules	JavaScript Course Modules
1	First contact with the Python language	Introduction to programming in Javascript
2	Character strings	Setting up programming environment
3	Programs	The Hello World! Program
4	Sequences	Variables
5	Loops	Data types
6	Sets and dictionaries	Comments
7	Functions	Operators
8	Object-oriented programming	Interaction with the user and dialog boxes
9	Python standard modules—overview	Conditional execution
10	Data processing	Loops
11	Algorithms in Python	Functions
12	Storage of data	Errors and exceptions
13	Use of PYPI modules	Testing your code
14	Python in practical applications	Cross-sectional task

It is worth mentioning that modules 1–9 cover the introductory Python curriculum, whereas modules 10–15 were included to provide students with a glimpse of Python programming practice, in the form of solving typical computing tasks using Python and its

libraries, and to make the students aware of what they need to learn to become capable of dealing with more complex tasks.

The basic form of content presentation in the course is textual (using different text formats), sometimes with illustrations. There are also code fragments intended to be executed in the Python interactive mode. One hundred and two course units selected as crucial for achieving the learning objectives are enriched with instructional videos. The audio track of each of these films corresponds to the basic text of a given unit and, possibly, parts of neighboring units, for which no separate films were recorded. All videos are provided with subtitles. A discussion forum is provided as an element of cooperative learning among the course participants.

Each lesson of the course includes at least one automatically graded exercise for the ongoing verification of the acquired knowledge. Exercises are either puzzles solved by dragging and dropping pieces of code, open questions, or multiple-choice closed questions. The final course grade is determined by the results obtained in the module summative tests and the closing test. The points acquired in these tests are treated equally, and to receive an electronic certificate of successful completion of the course, the course participant must achieve a threshold of at least 70% [50].

The “JavaScript Fundamentals” MOOC (later referred to as the JavaScript course) was designed as a self-study course with contents suitable for university and high school students, as well as self-learning programming enthusiasts. Like the Python course, the JavaScript course was also developed in two language versions—Polish and English—to make it available for English-speaking students. The course structure comprises modules (each ending with a test), lessons, and units (several units have the form of a lab, in which users are supposed to write code to solve a given case problem). Moreover, it includes an introduction with an opening test to assess the level of knowledge of the topic at the beginning of the course and the last module with the final exam to assess the level of knowledge at the end of the course. No elements enforcing the pace of learning were used—the only limitation was the general time frame of the course (its beginning and end date). The list of modules constituting the JavaScript course is presented in the third column of Table 2.

The basic form of content presentation in the course is textual (using different text formats), sometimes with illustrations. There are also code fragments that are intended to be executed in the JavaScript interactive mode. The course units are enriched with summary videos. All videos are provided with subtitles. A discussion forum is provided as an element of cooperative learning among the course participants.

Some lessons during the course include practicing the acquired knowledge, in which the learner is asked to write or correct a piece of code in JavaScript. Each module (except the last one, which is a cross-sectoral task) ends with a test that is graded automatically. These tests include single-choice and multiple-choice closed questions. The final course grade is determined by the results obtained in the module tests (20%) and the final exam (80%). To receive an electronic certificate of successful completion of the course, the course participant must achieve a threshold of at least 51%.

It should be noted that course elements used in both courses are fully in line with the results of Wong’s research on the factors influencing course completion, respectively, with regard to [51]:

- Encouragement to learn (detailed introduction—indicated by 100% of Wong’s respondents).
- Engagement (availability of multimedia—97% indications).
- Online interaction (discussion forum—100% indications).
- Consolidation of knowledge (automatically graded tests—81% indications).

To join any of the described courses, the platform user had to fill in a registration form, which included, among others, the following information: gender, age, education level, exact residential address (which indicated the city size), employment status, student status,

foreigner status, and disability status. For the purpose of this study, the participants' data have been anonymized.

2.2. Course Participants

Figure 1 presents demographic data regarding Python course participants ($n = 793$). As can be seen, male participants dominated (61%), and 58% of all course users were 25 or more years old. As regards the level of education, over 60% indicated higher education, while 12% selected that they were students. In terms of domicile, 43% admitted they live in a city with over 100 thousand inhabitants, and 87% of all learners were Poles. The remaining 13% included foreigners, as well as those who refused to answer this question. Unemployment status was denoted by 32% of the registered and 11% declared themselves as disabled (this group includes those who refused to provide this information). Finally, 86% of all registered users dropped the course and did not receive a certificate.



Figure 1. Characteristics of the Python course participants.

Figure 2 presents demographic data regarding JavaScript course participants ($n = 792$). As can be seen, in this course, male participants dominated (57%), and 55% of all course users were 25 years of age or older. As regards the level of education, 62% indicated that they had completed higher education, while 21% selected that they were students. In terms of domicile, 44% stated they lived in a city with over 100 thousand inhabitants, and 91% of all learners were Poles. The remaining 9% included foreigners, as well as those who refused to answer this question. Unemployment status was denoted by 43% of the registered and 10% declared being disabled (this group includes those who refused to provide this

information). Finally, 88% of all registered users dropped the course and did not receive a certificate.



* including those who refused to provide this information

Figure 2. Characteristics of the JavaScript course participants.

2.3. Research Procedure

The performed analysis was based on the data acquired from the course registration questionnaire, which, for both the considered courses, included: age, gender, education level, city, nationality, student status, unemployment status, and disability.

In the data-preprocessing step, all the considered variables were encoded as binary variables, as depicted in Table 3.

Table 3. Demographic variables considered in the model.

Variable	Description	Reference Value (0)
Gender	Participant’s gender	Female
Age	Participant’s age	Younger than 25 years old
Education	Participant’s higher education	No higher education
City	Participant’s place of living	City of less than 100,000 inhabitants
Student	Participant’s ongoing education	Not a student
Unemployed	Participant’s employment status	Employed
Foreigner	Participant’s country of origin	Poland
Disabilities	Participant’s disabilities	None

In order to analyze the relationship between demographic data of the Python course participants and the probability of completing the course, logistic regression was used [52]. Logistic regression is one of the most popular approaches to this kind of predictive problem, and has been successfully used for dropout prediction in the past [53,54]. A significance level of 0.05 has been assumed in the model.

All the data processing and calculations were performed using Python 3.8.2 and its pandas and statsmodels libraries.

3. Results

3.1. Zero-Time Prediction of Python MOOC Dropouts

The logistic regression model for the Python course dropout based on the eight demographic variables has a LLR p -value of less than 0.00001 which means it fits the data better than the null model. The model explains 10.7% of the dependent variable's variance. While this looks like a small number, considering it is based on zero-time predictors only, it actually supports the notion that the demographic predictors are useful for their intended purpose. Note that a result of 100% would mean that we could know if a person completes or drops out of the course based merely on their eight demographic descriptors, which would be an obvious absurdity.

In Table 4, the detailed prediction results obtained for the Python course are reported. Note that the table lists coefficients of the logistic regression model as the odds ratios, whereas in the text below it, we interpret the probabilities calculated from these results using the formula: $e^x / (1 + e^x)$.

Table 4. Results for the Python course.

	Coef	Std Err	z	P > z	[0.025	0.975]
const	−1.3583	0.309	−4.398	0.000	−1.964	−0.753
Gender	0.4457	0.235	1.899	0.058	−0.014	0.906
Age	−0.7256	0.256	−2.833	0.005	−1.228	−0.224
Education	−1.0213	0.265	−3.859	0.000	−1.540	−0.503
City	0.3790	0.216	1.759	0.079	−0.043	0.801
Student	−1.1476	0.385	−2.978	0.003	−1.903	−0.392
Unemployed	0.1680	0.273	0.616	0.538	−0.367	0.703
Foreigner	0.7724	0.347	2.226	0.026	0.092	1.453
Disabilities	−0.8490	0.424	−2.001	0.045	−1.681	−0.017

The model indicates that the probability of completing the Python course by a participant who is a female, under 25 years of age, with no higher education, who is not a student, is employed and not disabled, hails from Poland, and lives in a city with less than 100 thousand inhabitants equals to 20.45%.

The obtained results show that holding all other predictor variables constant, the probability of course completion decreased by 32.62% for users 25 years of age or older compared to younger users. Also, those who reported a higher education level were 26.48% less likely to complete the course compared to learners with lower education levels. Declaration of being a student also had a negative correlation with the probability of obtaining the course certificate (a 24.09% decrease in completion probability). A similar negative impact was observed for the disabled participants (a 29.96% decrease). The only predictor variable increasing the odds of course completion was being of a non-Polish nationality (which led to a 68.40% increase).

3.2. Zero-Time Prediction of JavaScript MOOC Drop-Outs

The logistic regression model for the JavaScript course dropout based on the eight demographic variables has a LLR p -value of 0.06021, which means it does not fit the

data better than the null model. Moreover, it explains merely 2.5% of the dependent variable's variance.

In Table 5, the detailed prediction results obtained for the JavaScript course are reported. The established threshold for statistical significance (0.05) has been met for none of the considered variables. These results mean the demographic variables are incapable of predicting the probability of the JavaScript course completion.

Table 5. Results for the JavaScript course.

	Coef	Std Err	z	P > z	[0.025	0.975]
const	−1.9642	0.337	−5.821	0.000	−2.626	−1.303
Gender	0.2847	0.226	1.258	0.208	−0.159	0.728
Age	0.0417	0.251	0.166	0.868	−0.449	0.533
Education	0.0303	0.287	0.105	0.916	−0.532	0.592
City	0.0455	0.220	0.206	0.837	−0.386	0.477
Student	−0.6583	0.408	−1.613	0.107	−1.458	0.141
Unemployed	−0.1440	0.280	−0.515	0.607	−0.692	0.404
Foreigner	−1.2818	0.668	−1.919	0.055	−2.591	0.027
Disabilities	0.1417	0.465	0.305	0.761	−0.770	1.054

4. Discussion

The results obtained for the Python course confirm that demographic variables can be effective predictors of MOOC dropout, though not all of them. Moreover, the results obtained for the JavaScript course show that this does not apply to all MOOCs, so their usability as zero-time predictors should be decided on a per-course basis. While we do not have sufficient background data to establish all reasons for the difference in results between the two analyzed courses, we are aware of one particular distinction: the Python course allowed only one answer per test question, whereas the JavaScript course permitted students to attempt the test as many times as they wished. Therefore, the Python course completion depended on both the students' persistence in continuing the course and their ability to comprehend the acquired knowledge, whereas the JavaScript course completion depended only on the former. This observation alone allows us to conclude that the prediction based on demographic data has been found to be more effective in pointing at the students being more susceptible to failing in learning programming than to those losing the will to continue the course. No similar (or adverse) observation has been found in the literature.

Looking at the respective demographic variables considered, our results—in contrast to some prior works [36,46], but in line with [3,16,28,29,37,45,47]—show that age could be used as a predictor of course completion or dropout. In our case, however, older participants (25 or more years old) were less likely to complete the course than younger learners, which is in line with [28,29] but contradictory to the remaining studies, which indicate older students are more successful [3,37,45,47].

The current work also supports most of the previous findings regarding gender, indicating that this variable has no significant impact on the completion/dropout rate [36,45–47,49], although studies showing different results also exist. In [3] authors claim that female participants are more likely to drop science courses, while the results of [29,37] suggest that females are generally more likely to drop out, no matter the course topic.

Our findings concerning the level of education support those presented in [28,37,45,48], showing that this feature can also be used as a course completion predictor but, in contrast to those reports, our case indicates that higher education of participants decreases the completion probability. There are also studies which have found that this variable is not significant in relation to dropout [36,46].

The only findings regarding the relationship of disability with completion [28] are similar to ours, indicating that disability negatively correlates with course completion.

Our findings regarding employment status concur with one of the studies [45], in which it was found that those who are not working are more likely to complete the course, but are in line with [47], who stated that employment has not been found to be significant in predicting the probability of learners' MOOC completion.

5. Conclusions

The presented results obtained for the first course (Python MOOC) clearly indicate that demographic data can be useful for predicting students' dropouts in MOOCs. However, by performing the same analysis for another course (JavaScript MOOC), we have demonstrated that this is not always the case. In line with the lack of compatibility of results reported in prior works, this allows us to draw a conclusion at a higher level of abstraction that the usefulness of demographic data for MOOC dropout prediction can and should only be determined on a per-course basis. Another higher-level conclusion is that this also applies to the set of demographic indicators that should be included in the prediction model, as the presented differences both between our two courses and with the other authors' results imply that there is not a single demographic variable that could be considered a reliable dropout predictor for all courses.

The presented study does, of course, have its limitations. Only two MOOCs were covered, which were similar in topic and their target groups, and only data from 1585 students were available for analysis. Nonetheless, given the character of the study outcomes (indicating the differences between the results reported for two analyzed courses rather than similarities), these limitations do not negate the value of the obtained results.

As for the practical implications of the presented results, the first is that demographic data should be considered in the models predicting students' dropout in MOOCs. This is an important indication, considering that these data are available from the moment at which a student registers for the course, unlike the behavior-based indicators (which can be measured only after some period of user activity) or other indicators that require third-party sources that may not always be available. Secondly, the decision on which demographic indicators to follow can only be made after some data are available and the relationship between specific variables and dropout has been confirmed or not. Note that this does not deprive the demographic indicators of their zero-time predictive ability for students newly joining the course; it just means that the set of effective indicators for a given course can only be determined after some students have participated in the course for some period of time.

The results of our study obviously indicate the direction of further research, which should determine why certain demographic variables have predictive power in some courses, but not in others. This line of research could presumably lead to obtaining some meta-model of dropout prediction, suggesting relevant prediction indicators based on other indicators.

Author Contributions: Conceptualization, J.S. and K.M.; methodology, J.S.; software, J.S.; validation, J.S.; formal analysis, J.S. and K.M.; investigation, K.M.; resources, K.M.; data curation, J.S.; writing—original draft preparation, J.S. and K.M.; writing—review and editing, J.S. and K.M.; visualization, K.M.; supervision, J.S.; project administration, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: The study is based on data collected for the primary purpose of course registration. The course participants were informed that these data could be used for analytical and research purposes.

Data Availability Statement: The anonymized data used in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. European Commission. Report on Web Skills Survey: Support Services to Foster Web Talent in Europe by Encouraging the Use of MOOCs Focused on Web Talent—First Interim Report. Available online: <https://silو.tips/download/report-on-web-skills-survey> (accessed on 15 October 2023).
2. Xie, X.; Siau, K.; Nah, F.F.-H. COVID-19 pandemic—Online education in the new normal and the next normal. *J. Inf. Technol. Case Appl. Res.* **2020**, *22*, 175–187. [CrossRef]
3. Feng, W.; Tang, J.; Liu, T.X. Understanding dropouts in MOOCs. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Association for the Advancement of Artificial Intelligence: Palo Alto, CA, USA, 2019; pp. 517–524.
4. Shah, D. The Second Year of the MOOC: 2020 Saw a Rush to Large-Scale Online Courses. Available online: <https://www.edsurge.com/news/2020-12-23-the-second-year-of-the-mooc-2020-saw-a-rush-to-large-scale-online-courses> (accessed on 15 October 2023).
5. Wang, M.; Li, G. A Meta-Analysis of MOOC-Based Academic Achievement, Engagement, Motivation, and Self-Regulation During the COVID-19 Pandemic. *Int. J. e-Collab.* **2022**, *18*, 1–17. [CrossRef]
6. Rõõm, M.; Lepp, M.; Luik, P. Dropout Time and Learners' Performance in Computer Programming MOOCs. *Educ. Sci.* **2021**, *11*, 643. [CrossRef]
7. Eriksson, T.; Adawi, T.; Stöhr, C. "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. *J. Comput. High. Educ.* **2017**, *29*, 133–146. [CrossRef]
8. de Freitas, S.I.; Morgan, J.; Gibson, D. Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. *Br. J. Educ. Technol.* **2015**, *46*, 455–471. [CrossRef]
9. Perna, L.W.; Ruby, A.; Boruch, R.F.; Wang, N.; Scull, J.; Ahmad, S.; Evans, C. Moving Through MOOCs: Understanding the Progression of Users in Massive Open Online Courses. *Educ. Res.* **2014**, *43*, 421–432. [CrossRef]
10. Goopio, J.; Cheung, C. The MOOC dropout phenomenon and retention strategies. *J. Teach. Travel Tour.* **2021**, *21*, 177–197. [CrossRef]
11. Şahin, M. A comparative analysis of dropout prediction in massive open online courses. *Arab. J. Sci. Eng.* **2021**, *46*, 1845–1861. [CrossRef]
12. Drousiotis, E.; Pentaliotis, P.; Shi, L.; Cristea, A.I. Capturing fairness and uncertainty in student dropout prediction—A comparison study. In Proceedings of the International Conference on Artificial Intelligence in Education, Utrecht, The Netherlands, 14–18 June 2021; Springer: Cham, Switzerland, 2021; pp. 139–144.
13. Fu, Q.; Gao, Z.; Zhou, J.; Zheng, Y. CLSA: A novel deep learning model for MOOC dropout prediction. *Comput. Electr. Eng.* **2021**, *94*, 107315. [CrossRef]
14. Mubarak, A.A.; Cao, H.; Hezam, I.M. Deep analytic model for student dropout prediction in massive open online courses. *Comput. Electr. Eng.* **2021**, *93*, 107271. [CrossRef]
15. Whitehill, J.; Mohan, K.; Seaton, D.T.; Rosen, Y.; Tingley, D. Delving Deeper into MOOC Student Dropout Prediction. *arXiv* **2017**, arXiv:1702.06404.
16. Bukralia, R. Predicting dropout in online courses: Comparison of classification techniques. In Proceedings of the Fifth Midwest Association for Information Systems Conference, Moorhead, MN, USA, 21–22 May 2010.
17. Lee, Y.; Choi, J. A review of online course dropout research: Implications for practice and future research. *Educ. Technol. Res. Dev.* **2011**, *59*, 593–618. [CrossRef]
18. Zhang, J.; Gao, M.; Zhang, J. The learning behaviours of dropouts in MOOCs: A collective attention network perspective. *Comput. Educ.* **2021**, *167*, 104189. [CrossRef]
19. Balakrishnan, G.; Coetzee, D. Predicting Student Retention in Massive Open Online Courses Using Hidden Markov Models. Available online: https://bid.berkeley.edu/cs294-1-spring13/images/7/7b/Balakrishnan%2C_Coetzee_-_Predicting_Student_Retention_in_Massive_Open_Online_Courses_using_Hidden_Markov_Models_-_CS294-1_project_report.pdf (accessed on 15 October 2023).
20. Kloft, M.; Stiehler, F.; Zheng, Z.; Pinkwart, N. Predicting MOOC dropout over weeks using machine learning methods. In Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs, Doha, Qatar, 25–29 October 2014; pp. 60–65.
21. Gregori, E.B.; Zhang, J.; Galván-Fernández, C.; de Asís Fernández-Navarro, F. Learner support in MOOCs: Identifying variables linked to completion. *Comput. Educ.* **2018**, *122*, 153–168. [CrossRef]
22. Jedel, I.A.V.; Palmquist, A.; Munkvold, R.I.; Goethe, O.; Jonasdottir, H.; Olsson, E.M. An industry experiment of academic performance and drop-out in gamified distance education. In Proceedings of the 5th International GamiFIN Conference 2021 (GamiFIN 2021), Levi, Finland, 7–9 April 2021.
23. Cristea, A.I.; Alamri, A.; Kayama, M.; Stewart, C.; Alsheri, M.; Shi, L. Earliest predictor of dropout in MOOCs: A longitudinal study of FutureLearn courses. In Proceedings of the 27th International Conference on Information Systems Development, Lund, Sweden, 22–24 August 2018.
24. Jordan, K. Massive open online course completion rates revisited: Assessment, length and attrition. *Int. Rev. Res. Open Distrib. Learn.* **2015**, *16*, 341–358. [CrossRef]

25. Rosé, C.P.; Carlson, R.; Yang, D.; Wen, M.; Resnick, L.; Goldman, P.; Sherer, J. Social factors that contribute to attrition in MOOCs. In Proceedings of the First ACM Conference on Learning @ Scale, Atlanta, GA, USA, 4–5 March 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 197–198.
26. Riofrío Calderón, G.; Ramírez Montoya, M.S.; Rodríguez Conde, M.J. Data analytics to predict dropout in a MOOC course on energy sustainability. In Proceedings of the 9th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2021), Barcelona, Spain, 27–29 October 2021.
27. Panagiotakopoulos, T.; Kotsiantis, S.; Kostopoulos, G.; Iatrellis, O.; Kameas, A. Early dropout prediction in MOOCs through supervised learning and hyperparameter optimization. *Electronics* **2021**, *10*, 1701. [\[CrossRef\]](#)
28. Radovanović, S.; Delibašić, B.; Suknović, M. Predicting dropout in online learning environments. *Comput. Sci. Inf. Syst.* **2021**, *18*, 957–978. [\[CrossRef\]](#)
29. Niemi, D.; Gitin, E. Using Big Data to Predict Student Dropouts: Technology Affordances for Research. In Proceedings of the International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2012), Madrid, Spain, 19–21 October 2012.
30. Adnan, M.; Habib, A.; Ashraf, J.; Mussadiq, S.; Raza, A.A.; Abid, M.; Bashir, M.; Khan, S.U. Predicting at-risk students at different percentages of course length for early intervention using machine learning models. *IEEE Access* **2021**, *9*, 7519–7539. [\[CrossRef\]](#)
31. Basnet, R.B.; Johnson, C.; Doleck, T. Dropout prediction in Moocs using deep learning and machine learning. *Educ. Inf. Technol.* **2022**, *27*, 11499–11513. [\[CrossRef\]](#)
32. Jha, N.I.; Ghergulescu, I.; Moldovan, A.N. OULAD MOOC Dropout and Result Prediction using Ensemble, Deep Learning and Regression Techniques. In Proceedings of the 11th International Conference on Computer Supported Education (CSEDU 2019), Heraklion, Greece, 2–4 May 2019; Scitepress: Setubal, Portugal, 2019; Volume 2, pp. 154–164.
33. Al-Shabandar, R.; Hussain, A.J.; Liatsis, P.; Keight, R. Detecting at-risk students with early interventions using machine learning techniques. *IEEE Access* **2019**, *7*, 149464–149478. [\[CrossRef\]](#)
34. Al-Shabandar, R.; Hussain, A.; Laws, A.; Keight, R.; Lunn, J.; Radi, N. Machine learning approaches to predict learning outcomes in Massive open online courses. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 713–720.
35. Kabathova, J.; Drlik, M. Towards predicting student's dropout in university courses using different machine learning techniques. *Appl. Sci.* **2021**, *11*, 3130. [\[CrossRef\]](#)
36. Hone, K.S.; El Said, G.R. Exploring the factors affecting MOOC retention: A survey study. *Comput. Educ.* **2016**, *98*, 157–168. [\[CrossRef\]](#)
37. Kizilcec, R.F.; Halawa, S. Attrition and achievement gaps in online learning. In Proceedings of the Second (2015) ACM Conference on Learning @ Scale (L@S '15), Vancouver, BC, Canada, 14–18 March 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 57–66.
38. Psathas, G.; Chatzidaki, T.K.; Demetriadis, S.N. Predictive Modeling of Student Dropout in MOOCs and Self-Regulated Learning. *Computers* **2023**, *12*, 194. [\[CrossRef\]](#)
39. Alhothali, A.; Albsisi, M.; Assalahi, H.; Aldosemani, T. Predicting student outcomes in online courses using machine learning techniques: A review. *Sustainability* **2022**, *14*, 6199. [\[CrossRef\]](#)
40. Pierrakeas, C.; Xeno, M.; Panagiotakopoulos, C.; Vergidis, D. A comparative study of dropout rates and causes for two different education courses. *Int. Rev. Res. Open Distrib. Learn.* **2004**, *5*, 183. [\[CrossRef\]](#)
41. Packham, G.; Jones, P.; Miller, C.; Thomas, B. E-learning and retention: Key factors influencing student withdrawal. *Educ. Train.* **2004**, *46*, 335–342. [\[CrossRef\]](#)
42. Tello, S.F. An analysis of student persistence in online education. In *Information Communication Technologies: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2008; pp. 1163–1178.
43. Levy, Y. Comparing dropouts and persistence in e-learning courses. *Comput. Educ.* **2007**, *48*, 185–204. [\[CrossRef\]](#)
44. Willging, P.A.; Johnson, S.D. Factors that influence students' decision to dropout of online courses. *J. Asynchronous Learn. Netw.* **2009**, *13*, 115–127.
45. Morris, N.P.; Swinnerton, B.; Hotchkiss, S. Can demographic information predict MOOC learner outcomes? In *Experience Track, Proceedings of the European MOOC Stakeholder Summit 2015, Mons, Belgium, 18–20 May 2015*; Université Catholique de Louvain: Ottignies-Louvain-la-Neuve, Belgium, 2015.
46. Sherimon, V.; Francis, L.; Pc, S.; Aboraya, W. Exploring the impact of learners' demographic characteristics on course completion and dropout in massive open online courses. *Int. J. Res. Granthaalayah* **2022**, *10*, 149–160. [\[CrossRef\]](#)
47. Zhang, Q.; Bonafini, F.C.; Lockee, B.B.; Jablokow, K.W.; Hu, X. Exploring demographics and students' motivation as predictors of completion of a massive open online course. *Int. Rev. Res. Open Distrib. Learn.* **2019**, *20*, 140–161. [\[CrossRef\]](#)
48. Rööm, M.; Luik, P.; Lepp, M. Learner success and the factors influencing it in computer programming MOOC. *Educ. Inf. Technol.* **2023**, *28*, 8645–8663. [\[CrossRef\]](#)
49. Breslow, L.; Pritchard, D.E.; DeBoer, J.; Stump, G.S.; Ho, A.D.; Seaton, D.T. Studying learning in the worldwide classroom research into edX's first MOOC. *Res. Pract. Assess.* **2013**, *8*, 13–25.
50. Swacha, J. Teaching Python programming with a MOOC: Course design and evaluation. In Proceedings of the Thirty-Seventh Information Systems Education Conference, Chicago, IL, USA, 9 October 2021; pp. 131–137.
51. Wong, B.T.-M. Factors leading to effective teaching of MOOCs. *Asian Assoc. Open Univ. J.* **2016**, *11*, 105–118. [\[CrossRef\]](#)

52. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
53. He, J.; Bailey, J.; Rubinstein, B.; Zhang, R. Identifying at-risk students in massive open online courses. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 1749–1755.
54. Jiang, S.; Williams, A.; Schenke, K.; Warschauer, M.; O’Dowd, D. Predicting MOOC performance with week 1 behavior. In Proceedings of the 7th International Conference on Educational Data Mining, London, UK, 4–7 July 2014.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.