Proceedings of the 56th Hawaii International Conference on System Sciences | 2023

A systematic review of the factors that impact the prediction of retention and dropout in higher education

Edmilson Cosme da Silva
Dept. of Computer Science
University of Brasília - Brazil
edmilsoncosme@unb.br

Sergio Antonio Andrade de Freitas
Faculty of Gama - FGA/UnB
University of Brasília - Brazil
sergiofreitas@unb.br

Cristiane Soares Ramos
Faculty of Gama- FGA/UnB
University of Brasília - Brazil
cristianesramos@unb.br

Amanda Emilly Muniz de Menezes
Faculty of Gama - FGA/UnB
University of Brasília - Brazil
amanda.muniiz@outlook.com

Letícia Karla S. Rodrigues de Araújo
Faculty of Gama - FGA/UnB
University of Brasília - Brazil
Leticiards.04@gmail.com

Abstract

Identifying factors that affect academic dropout and retention is a research area that brings a plurality of opinions and concepts. This article identifies current primary studies to understand the main factors related to dropout and retention. It is quantitative, exploratory, and explanatory research of an applied nature, using the technical procedures of case study and bibliographic research. The systematic review of the literature identifies the factors that impact academic dropout and retention and serves as a basis for a machine learning project. Academic, demographic, and learning factors can predict dropouts and retention. The definition of the factors used and the way of use is essential to obtain good forecasting results. The identified factors were used in the institution.

Keywords: Higher education management, Academic dropout and retention, Prediction, Factors, Systematic review.

1. Introduction

The education environment is highly complex when considering the student's characteristics and behaviors during his undergraduate course. Each student produces many data types during his academic life (ElAtia et al., 2016).

Education in higher education is responsible for providing a quality education for the student. Higher education institutions (HEIs) that generate good professionals are respected. However, most HEIs suffer from academic dropouts, which negatively impacts universities. The challenge of achieving student permanence in universities leads to a significant study of the factors that trigger dropout, such as sociodemographics, learning difficulties, and academic performance (Sadati and Libre, 2017).

The academic dropout theme generates curiosity in researchers, especially those focused on education. Mainly because of the impact on HEIs and students (Pereira et al., 2014). Many researchers try to link technology and the factors that lead to dropout to predict and avoid it.

The main objective of this study is to list the factors used by researchers to identify dropout and retention and how these factors can be found and classified in management systems and educational learning. With the application of the SRL, 52 studies related to retention and dropout were analyzed, and data from the factors presented in them were collected.

This paper is structured as follows: Section 2 presents the context of the University of Brasilia, as well as the definitions of retention and dropout used in the present study. Section 3 contains the protocol of systematic review. Section 4 describes the results and makes an analysis. Section 5 discusses the findings, and finally, in section 6, the conclusions are presented.

2. The research background

2.1. The academic context

Institutional managers and course coordinators focus on understanding student success behavior and developing the ability to predict the characteristics of students to cope with dropout and retention. Research shows that early identification and intervention are vital aspects that can lead to student graduation (Raju and Schumacker, 2015).

URI: https://hdl.handle.net/10125/102787
978-0-9981331-6-4

2.2. Dropout

Research on dropouts in higher education courses has been the subject of studies by researchers for some time now (Rumberger and Lim, 2008; Tinto, 1975). Two different perspectives can describe dropouts.

The foremost contemplates the student who drops out of university, and the second refers to the student who gives up higher education (Tinto, 1975). The first perspective harms the institution and causes losses related to reputation, income, and the opportunity to contribute to the student's life. In the second viewpoint, the lost workforce directly affects society, failing to rely on people qualified for the global market (Atif et al., 2013).

Studies in Brazil started to become more recurring in 1995, with the creation of the Special Commission for Studies on Dropout in Brazilian Public Universities (ANDIFES, 1996). This report brings together a set of essential data on the performance of Brazilian public universities. These data include students' graduation, retention, and dropout rates in their undergraduate courses. The commission makes a distinction of the term dropout into types:

- Course drops out: when the student withdraws from the higher education course in different situations such as abandonment (fails to enroll), withdrawal (official), transfer or re-option (change of course), exclusion by institutional rule,
- Dropout from the institution: when the student enrolled in an institution decides to leave,
- System dropout: the student leaves any form of higher education.

2.3. Retention

Some researchers say retention occurs when students complete, continue or resume their studies . Student retention/success happens when graduation is achieved (Raju and Schumacker, 2015).

Based on these studies, it is possible to identify a positive connotation in understanding the term retention, being applied in the sense of success in achieving some objective.

In Brazil, the report prepared by ANDIFES (ANDIFES, 1996) defines retention as the permanence in courses beyond the maximum time of curricular completion. The term retained is described as the condition in which the student, despite the expiry of the maximum period of curricular integration established by the Federal Council of Education (CFE), currently the National Council of Education (CNE), has not yet completed the course, remaining, however, enrolled at the university.

Based on the national surveys analyzed, it is possible to perceive a negative connotation for the term retention. The definitions used relates to the concept of time extrapolation defined by the curriculum (ANDIFES, 1996). Retention is secondary, given that most of it focuses on dropout. However, these concerns must often be approached as analogous phenomena or as a cause of each other (Pereira et al., 2014).

2.4. Related work

Silva and Souza's work (Silva et al., 2020) aims to identify through a Systematic Mapping of Literature the approaches and predictive techniques used to predict educational problems in teaching-learning environments. They also identify the factors that affect the learning process.

Saa et al. (Abu Saa et al., 2019) studied factors that affect the student's performance and the data mining techniques that use these factors. They revised 36 research articles and found four categories of the factors: students' previous grades and class performance, students' e-Learning activity, students' demographics, and students' social information.

Oliveira et al. (De Oliveira et al., 2021) present a Systematic Literature Review (SLR) for identifying the use of Educational Data Mining methodologies, techniques, and tools for preventing retention in higher education. They focus on identifying the Machine learning methods and recommendation systems.

3. The systematic review protocol

According to Kitchenham et al., 2015, a Systematic Literature Review (SLR) is a way of identifying, analyzing, and interpreting available evidence related to a particular research question, area, or phenomenon of interest. It involves nine activities, grouped into three phases: Planning the Review, Conducting the Review, and Documenting the Review.

Planning the review: addresses how to do the study. In it, three crucial activities take place: specification of research questions, development of the review protocol, and validation of the protocol.

Conducting the review: aims to identify relevant articles. The activities in this phase comprehend the following: selection of primary studies, quality assessment, data extraction, and data synthesis. Then the definitions described in the protocol are executed, and any discrepancies where changes to the protocol are required must be documented.

Documenting the review: applies the registration of the process and results.

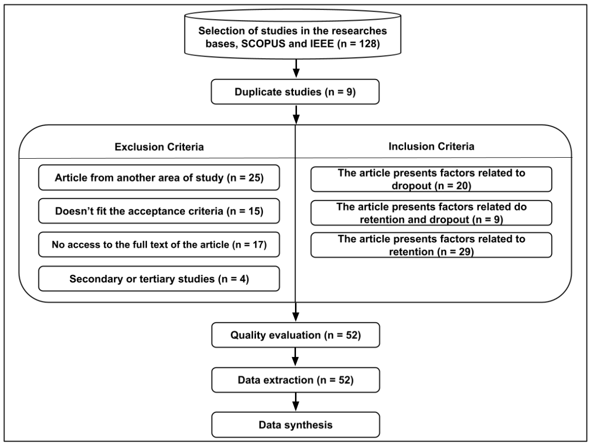

Figure 1. Overview of the systematic review process.

3.1. Planning the review

Research questions define the reasons and influence the entire review process (Kitchenham et al., 2015). For the SLR applied in this study, the questions are as follows:

- RQ1 - What factors are associated with the prediction of retention and dropout in higher education?
- RQ2 - What are the factors used to predict retention in higher education?

Once the questions were defined, they were documented in the protocol starting the development stage. This paper details the procedures performed during the SLR, which allows for review by other researchers, *feedback*, resolution of possible disagreements, and reduction of bias on the part of the researcher (Kitchenham et al., 2015). During the protocol development stage, the search expression considers the previously defined SLR objectives and the research questions. The search expression, once defined, was executed in SCOPUS digital libraries and the IEEE. Thus, the expression was defined as follows:
(student OR undergraduate) AND (predict* AND (dropout OR retention OR attrition)) AND (metric* OR measurement OR indicate*) AND (”higher education” OR ”bachelor degree”)) AND (LIMIT-TO (DOCTYPE, ”ar”) OR LIMIT-TO (DOCTYPE, ”cp”).
Other vital procedures are defined in the protocol development: study selection, quality assessment, data extraction, and data synthesis and aggregation. These procedures of the SLR conduction phase have their descriptions in the next section.
The protocol validation step plays a pivotal role in conducting a systematic review. Researchers must make agreements to evaluate the protocol. These agreements aid in generating strategies for evaluating the search expression and methods for extracting data (Kitchenham et al., 2015).

3.2. Conducting the review

This phase starts right after the protocol evaluation agreements, the following steps perform sequentially, but nothing prevents some from being performed simultaneously (Kitchenham et al., 2015).

In the Research Identification step, the search expression, defined in the Planning phase, was executed in the SCOPUS and IEEE databases. IEEE covers the essential software engineering and computer journals in general. SCOPUS, in turn, indexes, in a general way, papers published by ACM, Elsevier, Wiley, and Springer (Kitchenham et al., 2015). With this, it was possible to identify 128 articles, 118 in SCOPUS and 10 in IEEE.

The main objective of the Selection of Studies is to apply selection criteria to identify studies relevant to the research. Inclusion and exclusion criteria form the selection criteria. In this activity, on average, three blocks of 43 studies were created, contemplating the 128 articles. For each one of these blocks, a researcher was responsible for reading and applying the criteria through the analyses of the article's keywords and abstract. At the end of this process, the researcher evaluated the criteria used for inclusion and exclusion of the articles in the other two blocks assessed by the other researchers. The evaluation made possible the triangulation, whose purpose was to reinforce each other in providing evidence to validate the criteria for the inclusion or exclusion of the selected articles (Kitchenham et al., 2015). The product of this activity resulted in 52 articles being delivered for the quality assessment stage.

The next step is to Evaluate the Quality of the selected studies, where its objective is to determine if the empirical research is valid and unbiased (Kitchenham et al., 2015). At this stage, the researchers read the complete articles. No paper was removed, considering the small number of studies selected.

A form built for the Data Extraction stage was applied, where information regarding the factors that cause dropout and retention was cataloged.

Moreover, in the Data Synthesis stage, the information was tabulated consistently with the review questions, performing quantitative analyzes of the identified dropout and retention factors. Figure 1 details the activities performed during the conduct phase of the review.

3.3. Documenting the review

In this final phase of the SLR process, the study documentation adequately reports the study to the intended audience (Kitchenham et al., 2015).

4. Analysis of results

4.1. Dropout and Retention Factors

From the analyzed studies, it is possible to perceive that the causes that lead the student to leave higher education are the result of a combination of factors, with a degree of complexity and interconnectedness that evolve over the time of permanence (Atif et al., 2013).

In this context, it is possible to consider that retention and dropout are closely related. Both occur in the university environment involving the same actors (students, teachers, family members, colleagues, and others involved). Therefore, the factors that cause dropout applies to retention studies by observing the restrictions of each theory (Pereira et al., 2015).

The works selected in the SLR were evaluated by responding to RQ1 described in the planning phase of the systematic review, Section 3.1. This question aims to identify the factors involved in education retention and the dropout process.

The research advocates that the factors that lead to retention or dropout in higher education are related to the knowledge base cultivated before graduation — searching for information on undergraduate subjects to predict student performance in undergraduate courses (Esmat and Pitts, 2020; Tucker and McKnight, 2019). An example is considering high school grade point average (GPA) as having a significant correlation with university persistence (Raju and Schumacker, 2015).

Studies to evaluate the habits of students of technology courses based on the skills of studies deeply and superficially: in-depth, the student seeks to associate information with previously acquired knowledge; in superficial, the student tries to memorize new subjects without associating them with prior understanding (Atieh et al., 2020).

Researchers direct their studies associated with performance in subjects, exploring the relationship between students enrolled in STEM courses and their performance in subjects where the use of mathematics is strong (Cohen and Kelly, 2020).

Socio-demographic factors are significant predictors of academic success (Cano and Leonard, 2019). In student entrance procedures, for example, the inclusion of socio-economic context improves the examination process (Darabi et al., 2017).

An essential source for identifying factors related to the study of dropout and retention is directly related to educational systems. These systems are becoming a growing source of helpful information to predict student behavior (ElAtia et al., 2016; Sadati and Libre, 2017).

With the evaluation of the factors that cause dropout and retention using quantifiable formats and an understanding of the ways of grouping these factors, it is possible to use the term indicator to represent them in the context of this study. This representation enables objective measurement and better systematic representation in an educational context. These indicators are described in Section 4.2.

4.2. Prediction of retention and dropout in higher education

This section aims to respond to RQ2 described in the planning phase of the systematic review. The identification and classification of factors define the context of the use of each factor.

The identification resulted in 29 factors presented in Table 1. The ten most cited are average score, gender, course grades, degree, age, ethnicity, scholarship, zip code, regular study, course schedule, and university entrance type.

Retention is calculated based on defined rates, considering demographic and academic variables, initial and motivational aspirations of students, personality and value of students, and institutional and interaction variables with educational systems. This variety of variables makes calculating fees complex (Atif et al., 2013).

The factors listed in this study were categorized considering their characteristics, which resulted in three groups: demographic, academic, and learning.

The Figure 2 contains the eleven demographic factors, Figure 3 shows the eleven academic factors and Figure 4 contains the seven most used learning factors.

The student's average score was the most used factor in evaluating dropout and retention cases in higher education.

With the use of the average preparation course punctuation, it is possible to predict the scores of candidates for vacancies in engineering faculties (Al-Sheeb et al., 2019; Darabi et al., 2017). Research has considered assessing student retention in STEM courses based on average math scores and average

Table 1. Identified factors and references' index

Factor	Total	Reference (index at table 2)
average scores	27	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
gender	23	2, 3, 6, 7, 10, 11, 12, 14, 15, 16, 18, 23, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37
course grades	18	1, 3, 4, 6, 7, 12, 13, 15, 16, 21, 23, 24, 27, 31, 38, 39, 40, 41
course	17	1, 4, 7, 9, 10, 15, 18, 19, 20, 23, 26, 27, 34, 36, 37, 42, 43
age	14	3, 10, 11, 13, 14, 15, 16, 18, 30, 31, 32, 36, 37, 39
ethnicity	13	2, 3, 5, 6, 7, 23, 27, 28, 32, 33, 36, 43, 44
scholarship	12	3, 5, 10, 15, 24, 29, 33, 35, 36, 39, 42, 44
zip code	10	3, 10, 11, 12, 13, 23, 29, 33, 35, 45
regular study	8	6, 10, 16, 18, 26, 41, 44, 46
course timetable	8	2, 4, 22, 26, 36, 39, 44, 47
entrance type	8	12, 13, 15, 16, 37, 39, 45, 48
job	7	3, 11, 29, 30, 31, 33, 35
number of approved subjects	7	7, 15, 19, 44, 47, 49, 50
parents' educational level	6	3, 8, 11, 29, 32, 42
forum access	5	23, 31, 38, 41, 46
access number	5	22, 23, 38, 51
exercise solving	4	23, 38, 41, 46
video access	3	41, 46, 52
previous knowledge	3	13, 16, 31
marital status	3	10, 31, 36
foreigner	3	15, 27, 39
number of failed subjects	3	13, 39, 44
special needs	3	8, 22, 48
lock number	3	17, 44, 49
number of courses enrolled	2	11, 44
research data	2	11, 24
children	2	30, 31
belated submission	1	46
internship	1	44

grades for the first-semester (Al-Sheeb et al., 2019; Cohen and Kelly, 2020).

The factors of gender and ethnicity are significant predictors of retention at graduation (Raju and Schumacker, 2015).

Analyzing students' behavior at a university in Austria, it was possible to identify those male students who were fluent in German and had exemplary performance in high school had better academic success rates (Frischenschlager et al., 2005).

Assessing the use of electronic portfolios, balanced with other factors, can improve the prediction of dropouts. Measuring students' interaction with this system and gender-type factors showed correlation gains in the analyses (Aguiar et al., 2014).

The scores of standardized tests such as the SAT[1], can predict student performance (Ford et al., 2012; Tekin, 2014). In the article by Reed et al., 2012, the freshman success rate correlates to SAT scores above 1000 points.

The results of the Tharp, 1998 research showed that students enrolled in two-year courses had better persistence rates than undergraduate students. The hours and average points of the first semester presented a significant value in its ability to predict dropout.

In the Beck and Milligan, 2014 research, the degree of institutional commitment, where subjects such as loyalty, trust in selecting the institute, how much the student was incorporated or satisfied with the teaching conditions offered by the university, the correlation analyses showed that older students had higher institutional commitment values.

To identify factors related to the success of medical students perceived that maturity with the association of

[1] *Scholastic Aptitude Test* (SAT): https://www.studentprogress.org/gre/scholastic-aptitude-test-sat/

Table 2. Index of references

Idx	Reference	Idx	Reference	Idx	Reference
1	Sadati and Libre, 2017	19	Meens et al., 2018	37	Cardona et al., 2020
2	Raju and Schumacker, 2015	20	Vos et al., 2019	38	Figueroa and Sancho, 2020
3	Atif et al., 2013	21	Tekin, 2014	39	Oreshin et al., 2020
4	Esmat and Pitts, 2020	22	Deighton et al., 2019	40	Ford et al., 2012
5	Tucker and McKnight, 2019	23	Aguiar et al., 2014	41	Liu et al., 2015
6	Atieh et al., 2020	24	Eubanks et al., 2016	42	Fernández-Martíen et al., 2019
7	Cohen and Kelly, 2020	25	Wade, 2019	43	Luciano-Wong and Crowe, 2019
8	Cano and Leonard, 2019	26	Al-Sheeb et al., 2019	44	da Silva et al., 2019
9	Darabi et al., 2017	27	Reed et al., 2012	45	Luis Arroyo-Barriguete et al., 2020
10	Sani et al., 2020	28	Dey and Astin, 1993	46	Kondo et al., 2017
11	Frischenschlager et al., 2005	29	Fényes et al., 2021	47	Celis et al., 2019
12	Lázaro Alvarez et al., 2020	30	A. Pérez et al., 2018	48	A. M. Pérez et al., 2018
13	Baranyi et al., 2019	31	Kostopoulos et al., 2017	49	Bossema et al., 2017
14	Bargmann et al., 2021	32	Beck and Milligan, 2014	50	Adejo and Connolly, 2018
15	Kiss et al., 2019	33	Campbell and Mislevy, 2013	51	Klein et al., 2019
16	Kilian et al., 2020	34	Rintala et al., 2011	52	Respondek et al., 2017
17	Niessen et al., 2016	35	Hoffman and Lowitzki, 2005		
18	Reparaz et al., 2020	36	Tharp, 1998		

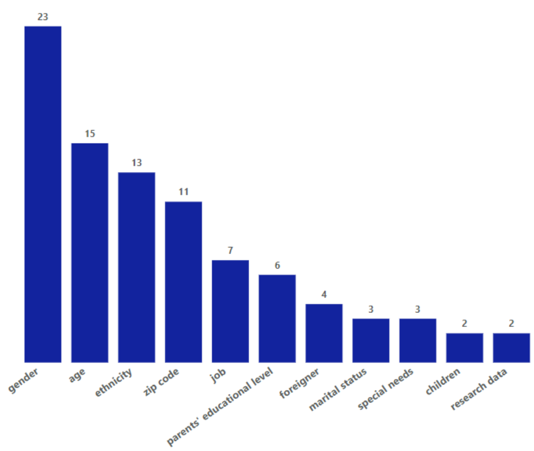

Figure 2. Demographic Factors

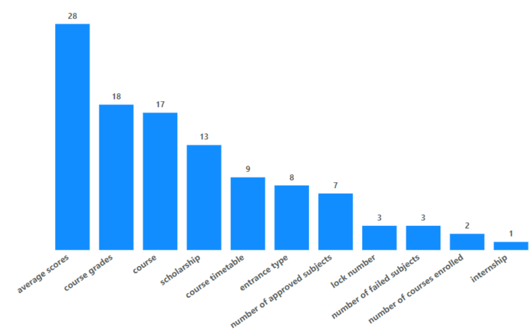

Figure 3. Academic Factors

other factors significantly influenced the success of these students (Frischenschlager et al., 2005).

The student's place of residence was one of the factors that played an essential role in Campbell and Mislevy, 2013's research. Women not living close to the university are more likely to transfer instead of continuing the course at the institution.

5. Discussion

5.1. Findings

This study aims, at first, to list factors used by researchers to identify dropout and retention and how these factors can be found and classified in management and educational learning systems.

In a second step, these factors will be worked on to make them indicators used in traditional statistical summaries to report descriptive information about operations, including enrollment and retention indicators, financial aid, revenue forecasts, and learning assessments (Eubanks et al., 2016).

The factors found can be used by institutes, where knowledge extracted from past and current datasets can represent and provide information to university administrators to monitor conditions and take measures to solve issues (Tekin, 2014).

Research by Hoffman and Lowitzki, 2005 has suggested that a significant body of publications has indicated that average high school scores (GPA) and scores on standardized tests such as the SAT are generally decisive predictors of academic success in

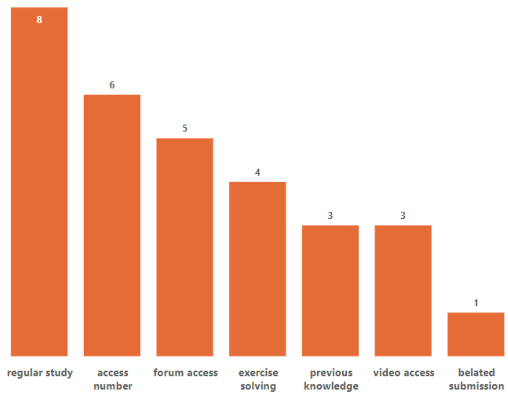

Figure 4. Learning Factors

college. This work confirms what was exposed by the researcher, demonstrating that the average score was first and the course grades third in the general result described in Table 1.

Separating factors into categories (demographic, academic learning) makes it possible to comprehend these factors' space in educational systems. Most studies attribute academic factors as the best predictors of dropout and retention (Niessen et al., 2016; Tekin, 2014). The analyses show that diversifying these factors can boost the predictive power of the systems developed to support this activity. Ethnicity combined with other data can be highly significant in the retention or dropout analysis (Raju and Schumacker, 2015).

The combination of demographic and learning factors taken from the Learning System or carried out in Course Management can present actual results. Given the broad implementation of these systems and their level of maturity, studies that involve analysis of the learning factors taken from these systems can also be part of studies that serve other educational institutions (Atif et al., 2013).

5.2. Undergoing research

The University of Brasilia (UnB) is a public university in Brazil. It has 134 active undergraduate courses and more than 40.000 undergraduate students This large number of students brings the university the challenge of keeping scholars engaged and focused on completing their chosen course.

UnB's Institutional Development Plan shows that the percentage of students who remained at the university after the expected conclusion time - retention rate - is about 50.16%, and the dropout rate is about 24.54%. The low-income students in the inclusive policies are about 40%.

The identified indicators of the present work are being mapped on the institution's data system, enabling the analysis of the students' dropout and retention. Institutional academic systems have data on students since 1990. With this data and through data mining and machine learning, a support system is being implemented for managers, course coordinators, and teachers for forecasting and decision-making to minimize university dropout and retention rates.

Thus, it is possible to point out, for example, which disciplines have the highest retention rate and the indicators that characterize the profiles of students during the academic semester for each of them.

Since 2017 UnB has had an institutional program to encourage active methodologies (A3M - Learning for the Third Millennium). Since its creation, more than 110 projects proposed by teachers who wish to innovate their teaching methods and improve learning indicators in their disciplines have been implemented. In this context, it is expected that we will be able to cross-reference the identified indicators with the learning indicators, contributing to the identification of the most effective teaching methods for reducing the retention and dropout rate.

6. Conclusion

The present systematic literature survey was a fundamental step in identifying the factors predicting dropout and retention. It allows a new perspective on the mass of data already existing in educational institutions. In particular, for the prevision of retention and dropout of students.

These factors enable the school to identify what information it has in its databases. Once identified, this information's availability allows us to plan a machine learning project.

There was two potential interference in the results obtained. The first one occurred in the execution of the search string in two digital libraries, which may have reduced the number of papers returned. In addition, there was challenging to compare the concept of retention with international definitions since there were differences between national systems, concepts, and performance indicators.

Future research will use the data found and select the best variables to construct models using data mining techniques such as regression, decision trees, and neural networks. The purpose is to find essential characteristics associated with dropout and retention.

References

Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, *24*. https://doi.org/10.1007/s10758-019-09408-7

Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*.

Aguiar, E., Chawla, N. V., Brockman, J., Ambrose, G. A., & Goodrich, V. (2014). Engagement vs performance: Using electronic portfolios to predict first semester engineering student retention. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 103–112.

Al-Sheeb, B. A., Hamouda, A., & Abdella, G. M. (2019). Modeling of student academic achievement in engineering education using cognitive and non-cognitive factors. *Journal of Applied Research in Higher Education*.

ANDIFES. (1996). Special commission for studies on dropout in brazilian public universities. Retrieved August 21, 2022, from https://www.andifes.org.br/wp-content/files_flutter/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf

Atieh, E. L., York, D. M., & Muñiz, M. N. (2020). Beneath the surface: An investigation of general chemistry students' study skills to predict course outcomes. *Journal of chemical education*, *98*(2), 281–292.

Atif, A., Richards, D., & Bilgin, A. (2013). A student retention model: Empirical, theoretical and pragmatic considerations.

Baranyi, M., Gál, K., Molontay, R., & Szabó, M. (2019). Modeling students' academic performance using bayesian networks. *2019 17th international conference on emerging eLearning technologies and applications (ICETA)*, 42–49.

Bargmann, C., Thiele, L., & Kauffeld, S. (2021). Motivation matters: Predicting students' career decidedness and intention to drop out after the first year in higher education. *Higher Education*, 1–17.

Beck, H. P., & Milligan, M. (2014). Factors influencing the institutional commitment of online students. *The Internet and Higher Education*, *20*, 51–56.

Bossema, E. R., Meijs, T. H., & Peters, J. W. (2017). Early predictors of study success in a dutch advanced nurse practitioner education program: A retrospective cohort study. *Nurse Education Today*, *57*, 68–73.

Campbell, C. M., & Mislevy, J. L. (2013). Student perceptions matter: Early signs of undergraduate student retention/attrition. *Journal of College Student Retention: Research, Theory & Practice*, *14*(4), 467–493.

Cano, A., & Leonard, J. D. (2019). Interpretable multiview early warning system adapted to underrepresented student populations. *IEEE Transactions on Learning Technologies*, *12*(2), 198–211.

Cardona, T. A., Cudney, E. A., Snyder, J., & Hoerl, R. W. (2020). Predicting student degree completion using random forest. *2020 ASEE Virtual Annual Conference Content Access*.

Celis, S., López, D., & Silva, J. (2019). Analyzing the influence of online behaviors and learning approaches on academic performance in first year engineering. *2nd Latin American Conference on Learning Analytics, LALA*, *2425*, 1–12.

Cohen, R., & Kelly, A. M. (2020). Mathematics as a factor in community college stem performance, persistence, and degree attainment. *Journal of Research in Science Teaching*, *57*(2), 279–307.

Darabi, H., Douzali, E., Harford, S. T., Nelson, P. C., & Karim, F. S. M. (2017). Beyond grade point average and standardized testing: Incorporating a socio-economic factor in admissions to support minority success. *2017 ASEE Annual Conference & Exposition*.

da Silva, P. M., Lima, M. N., Soares, W. L., Silva, I. R., Roberta, A. d. A., & de Souza, F. F. (2019). Ensemble regression models applied to dropout in higher education. *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, 120–125.

De Oliveira, T. N., Bernardini, F., & Viterbo, J. (2021). An overview on the use of educational data mining for constructing recommendation systems to mitigate retention in higher education. *2021 IEEE Frontiers in Education Conference (FIE)*, 1–7. https://doi.org/10.1109/FIE49875.2021.9637207

Deighton, K., Hudson, J., Manley, A. J., Kaiseler, M., Patterson, L. B., Rutherford, Z. H., & Swainson, M. (2019). Effects of emotional intelligence and supportive text messages on academic outcomes in first-year undergraduates. *Journal of Further and Higher Education*, *43*(4), 494–507.

Dey, E. L., & Astin, A. W. (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. *Research in higher education*, *34*(5), 569–581.

ElAtia, S., Ipperciel, D., & Zaıéane, O. R. (2016). *Data mining and learning analytics: Applications in educational research*. John Wiley & Sons.

Esmat, T. A., & Pitts, J. D. (2020). Predicting success in an undergraduate exercise science program using science-based admission courses. *Advances in physiology education*, *44*(2), 138–144.

Eubanks, D., Evers Jr, W., & Smith, N. (2016). Finding predictors in higher education. *Data Mining and Learning Analytics*, 41.

Fényes, H., Mohácsi, M., & Pallay, K. (2021). Career consciousness and commitment to graduation among higher education students in central and eastern europe. *Economics & Sociology*, *14*(1), 61–75.

Fernández-Martıén, T., Solıés-Salazar, M., Hernández-Jiménez, M. T., & Moreira-Mora, T. E. (2019). A multinomial and predictive analysis of factors associated with university dropout. *Revista Electrónica Educare*, *23*(1), 73–97.

Figueroa-Cañas, J., & Sancho-Vinuesa, T. (2020). Early prediction of dropout and final exam performance in an online statistics course. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, *15*(2), 86–94.

Ford, G. D., Steffen, R. E., & Ford, R. W. (2012). University freshman retention in north carolina. *2012 ASEE Annual Conference & Exposition*, 25–1400.

Frischenschlager, O., Haidinger, G., & Mitterauer, L. (2005). Factors associated with academic success at vienna medical school: Prospective survey. *Croatian medical journal*, *46*(1).

Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *The review of higher education*, *28*(4), 455–474.

Kilian, P., Loose, F., & Kelava, A. (2020). Predicting math student success in the initial phase of college with sparse information using approaches from statistical learning. *Frontiers in Education*, 248.

Kiss, B., Nagy, M., Molontay, R., & Csabay, B. (2019). Predicting dropout using high school and first-semester academic achievement measures. *2019 17th international conference on emerging eLearning technologies and applications (ICETA)*, 383–389.

Kitchenham, B. A., Budgen, D., & Brereton, P. (2015). *Evidence-based software engineering and systematic reviews* (Vol. 4). CRC press.

Klein, C., Lester, J., Rangwala, H., & Johri, A. (2019). Technological barriers and incentives to learning analytics adoption in higher education: Insights from users. *Journal of Computing in Higher Education*, *31*(3), 604–625.

Kondo, N., Okubo, M., & Hatanaka, T. (2017). Early detection of at-risk students using machine learning based on lms log data. *2017 6th IIAI international congress on advanced applied informatics (IIAI-AAI)*, 198–201.

Kostopoulos, G., Kotsiantis, S., Ragos, O., & Grapsa, T. N. (2017). Early dropout prediction in distance higher education using active learning. *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 1–6.

Lázaro Alvarez, N., Callejas Carrión, Z., Griol Barres, D., et al. (2020). Factores que inciden en la deserción estudiantil en carreras de perfil ingenierıéa informática. *Revista Fuentes*.

Liu, D., Richards, D., Froissard, C., & Atif, A. (2015). Validating the effectiveness of the moodle engagement analytics plugin to predict student academic performance. *21st Americas Conference on Information Systems, AMCIS 2015*, 1–10.

Luciano-Wong, S., & Crowe, D. (2019). Persistence and engagement among first-year hispanic students. *Journal for Multicultural Education*.

Luis Arroyo-Barriguete, J., Tirado, G., Mahillo-Fernandez, I., & Jose Ramirez, P. (2020). Predictors of performance in business administration degrees: The effect of the high-school specialty. *REVISTA DE EDUCACION*, (390), 125–148.

Meens, E. E., Bakx, A. W., Klimstra, T. A., & Denissen, J. J. (2018). The association of identity and motivation with students' academic achievement in higher education. *Learning and Individual Differences*, *64*, 54–70.

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Predicting performance in higher education using proximal predictors. *PloS one*, *11*(4), e0153663.

Oreshin, S., Filchenkov, A., Petrusha, P., Krasheninnikov, E., Panfilov, A., Glukhov, I., Kaliberda, Y., Masalskiy, D., Serdyukov, A., Kazakovtsev, V., et al. (2020). Implementing a machine learning approach to predicting students' academic outcomes. *2020 International Conference on Control, Robotics and Intelligent System*, 78–83.

Pereira, A. S., Carneiro, T. C. J., Brasil, G. H., & Corassa, M. A. d. C. (2014). Perfil dos alunos retidos dos cursos de graduação presencial da universidade federal do espıérito santo.

Pereira, A. S., Carneiro, T. C. J., Brasil, G. H., & Corassa, M. A. d. C. (2015). Fatores relevantes no processo de permanência prolongada de discentes nos cursos de graduação presencial: Um estudo na universidade federal do espıérito santo. *Ensaio: Avaliação e Polıéticas Públicas em Educação*, *23*, 1015–1039.

Pérez, A. M., Escobar, C. R., Toledo, M. R., Gutierrez, L. B., & Reyes, G. M. (2018). Prediction model of first-year student desertion at universidad bernardo o'higgins (ubo). *Educacao e Pesquisa*, *44*.

Pérez, A., Grandón, E. E., Caniupán, M., & Vargas, G. (2018). Comparative analysis of prediction techniques to determine student dropout: Logistic regression vs decision trees. *2018 37th International Conference of the Chilean Computer Science Society (SCCC)*, 1–8.

Raju, D., & Schumacker, R. (2015). Exploring student characteristics of retention that lead to graduation in higher education using data mining models. *Journal of College Student Retention: Research, Theory & Practice*, *16*(4), 563–591.

Reed, T. K., Imbrie, P., Jin, Q., & Lin, J. J. (2012). Modeling student success of international undergraduate engineers. *2012 ASEE Annual Conference & Exposition*, 25–945.

Reparaz, C., Aznárez-Sanado, M., & Mendoza, G. (2020). Self-regulation of learning and mooc retention. *Computers in Human Behavior*, *111*, 106423.

Respondek, L., Seufert, T., Stupnisky, R., & Nett, U. E. (2017). Perceived academic control and academic emotions predict undergraduate university student success: Examining effects on dropout intention and achievement. *Frontiers in psychology*, *8*, 243.

Rintala, U., Andersson, S., & Kairamo, A.-K. (2011). How to retain students in higher engineering education? findings of the attract project. *SEFI Annual Conference, Lisbon, Portugal. http://goo. gl/tdVMj5*.

Rumberger, R. W., & Lim, S. A. (2008). Why students drop out of school: A review of 25 years of research.

Sadati, S., & Libre, N. A. (2017). Development of an early alert system to predict students at risk of failing based on their early course activities. *2017 ASEE Annual Conference & Exposition*.

Sani, N. S., Nafuri, A. F. M., Othman, Z. A., Nazri, M. Z. A., & Nadiyah Mohamad, K. (2020). Drop-out prediction in higher education among b40 students. *International Journal of Advanced Computer Science and Applications*, *11*(11), 550–559.

Silva, P., Souza, F., & Fagundes, R. (2020). Approaches to predicting educational problems: A systematic mapping. *XVI Brazilian Symposium on Information Systems*. https://doi.org/10.1145/3411564.3411657

Tekin, A. (2014). Early prediction of students' grade point averages at graduation: A data mining approach. *Eurasian Journal of Educational Research*, *54*, 207–226.

Tharp, J. (1998). Predicting persistence of urban commuter campus students utilizing student background characteristics from enrollment data. *Community College Journal of Research and Practice*, *22*(3), 279–294.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, *45*(1), 89–125.

Tucker, L., & McKnight, O. (2019). Assessing the validity of college success indicators for the at-risk student: Toward developing a best-practice model. *Journal of College Student Retention: Research, Theory & Practice*, *21*(2), 166–183.

Vos, C. M., Wouters, A., Jonker, M., de Haan, M., Westerhof, M. A., Croiset, G., & Kusurkar, R. A. (2019). Bachelor completion and dropout rates of selected, rejected and lottery-admitted medical students in the netherlands. *BMC Medical Education*, *19*(1), 1–9.

Wade, N. L. (2019). Measuring, manipulating, and predicting student success: A 10-year assessment of carnegie r1 doctoral universities between 2004 and 2013. *Journal of College Student Retention: Research, Theory & Practice*, *21*(1), 119–141.