

PREDICTING COMPUTER ENGINEERING STUDENTS' DROPOUT IN CUBAN HIGHER EDUCATION WITH PRE-ENROLLMENT AND EARLY PERFORMANCE DATA

Niurys Lázaro Alvarez¹, Zoraida Callejas², David Griol²

¹Universidad de las Ciencias Informáticas (Cuba)

²University of Granada (Spain)

nlazaro@uci.cu, zoraida@ugr.es, dgriol@ugr.es

Received January 2020

Accepted May 2020

Abstract

We present an educational data analytics case study aimed at the early detection of potential dropout in Computer Engineering studies in Cuba. We have employed institutional data of 456 students and performed several experiments for predicting their permanency into three (promotion, repetition, and dropout) or two classes (promoting, not promoting). We have also tested a combination of classification features for training and testing decision trees and neural networks; including information obtained at the time of enrollment, after the first semester and after the first academic year. Our results show a considerable accuracy using all features (96.71%). Using only the features available at the time of enrolment and after the first semester we obtain very positive results (68.86% and 93.85% accuracy respectively) with a high recall of non-promoting students. Thus, it is possible to obtain an early assessment of the risk of dropout that can help defining prevention policies.

Keywords – Dropout, Retention, Promotion, Higher education, Data analysis, Computer engineering, Automatic classification.

To cite this article:

Lázaro Alvarez, N., Callejas, Z., & Griol, D. (2020). Predicting computer engineering students' dropout in Cuban higher education with pre-enrollment. *Journal of Technology and Science Education*, 10(2), 241-258. <https://doi.org/10.3926/jotse.922>

1. Introduction

The progress of a country is highly dependent on the education of its citizens (Munizaga-Mellado, Cifuentes-Orellana & Beltrán-Gabriel, 2018; Rodríguez-Gómez, Feixas, M., Gairín & Muñoz, 2012), as education plays a central role and has a cross cutting impact on all aspects of human life. Universities prepare students to perform as professionals and create environments for national advancement based on creativity and innovation, supporting economic growth for the country. Thus, it is fundamental to understand the causes for higher-education dropout in order to mitigate it. However, this is a complex task as dropout is a multifaceted phenomenon in which heterogeneous variables are implied.

Dropout in Higher Education institutions is becoming a key concern for educators and researchers. Students' dropout at the university level involves economic, social and psychological costs (Allen, Robbins, Casillas & Oh, 2008; Burgos, Campanario, Pena, Lara, Lizcano & Martinez, 2018; Freeman, Anderman & Jensen, 2007; Hausmann, Schofield & Woods, 2008; Lansford, Dodge, Pettit & Bates, 2016; Lotkowski, Robbins & Noeth, 2004; Ramsdal, Gurgens & Wynn, 2013).

Although several studies have been conducted to explore the factors associated with dropout and student retention, there are still disagreements. Several theories have been developed to understand why and how students decide to dropout. Economical, organizational, psychological and sociological perspectives have been considered by different approaches (Braxton & Hirschy, 2005; Gregori, Martínez & Moyano-Fernández, 2018; Itzhaki, Itzhaky & Yablon, 2018). In addition to the theoretical models, several studies are focused on discovering the relationships and correlations between university retention and attrition. In most studies about university dropout, demographic and academic variables have traditionally been used to predict the success and academic promotion of university students (Pritchard & Wilson, 2003; Reason, 2003; Robbins, Allen, Casillas, Peterson & Lee, 2006). Traditional academic variables, such as previous academic performance and educational development, have also been proved to have an indirect effect on university dropout (Allen et al., 2008).

The demand of graduates in the branches of Science, Technology, Engineering and Mathematics (STEM, for its acronym in English) has experienced an increase due to the growing scientific-technological development and its applications. However, recent statistics show that dropouts in these degrees are very high (Peña-Calvo, Inda-Caro, Rodríguez-Menéndez & Fernández-García, 2016). In 2015 there was a growth of 13% in the demand of engineers and scientists in Europe and it is foreseen that this demand will increase by 14% by 2025. However, this contrasts with the fact that the enrollment of students of Engineering and Architecture degrees decreased by 24.6% from 2004 to 2014 in the same context, which implies that one out of every four students have left these degrees (López, Carpeño, Arriaga & Ruiz, 2016).

In Latin America and the Caribbean several studies have been carried out describing the variables and factors that affect their different countries (Costa, Bispo & Pereira, 2018). This is the case of the studies of the Institute of Higher Education for Latin America and the Caribbean (IESALC) that pursued measuring the magnitude of university dropout in 15 countries and 3 areas of knowledge: Law, Medicine and Civil Engineering. (Munizaga-Mellado et al., 2018) contributed a systematic review of publications from 1990 to 2016, identifying among other aspects: countries of study, methodologies used, variables and factors identified. They analyzed 81 articles that responded to their objectives, of which only 3 are Cuban, 6 are dedicated to engineering degrees and only one specifically to Computer Engineering. Apart from these studies, Latin America is underrepresented in engineering education research (Williams, Wankat & Neto, 2018).

The Cuban Ministry of Higher Education periodically analyzes several quality indicators including retention and academic efficiency. Even though Cuba is one of the Latin American countries with lower dropout rates, for engineering careers it is around 50% (IESALC-UNESCO, 2007). Thus, our study addresses the problem of the desertion of higher education in the context of Computer Engineering profile degrees in Cuba, given the importance of these engineers for the growth of the country.

This study poses two research questions in the Cuban context:

- How accurately do Computer Engineering students' characteristics predict the risk of dropout?
- How accurately can dropout be predicted before the end of the first year?

In line with the related literature, our study addresses the role of traditional and non-traditional predictors of university students' dropout intentions. To the best of our knowledge, there are no previous studies addressing the problem of higher education dropout in the context of the Computer Engineering degrees in Cuba.

In addition, previous work is devoted mostly to reveal the multilayered structure of causes that explain dropout, but do not necessarily follow a practical manageability and applicability of the models. In order for the models to be used in practice, it is necessary to test them in particular settings and have the ability to feed them with student's data to generate predictions. With the dropout prediction approach followed in this study, it is possible to identify the students' dropout risk in the Cuban Computer Engineering setting. Such evidence will inform the systemic operation of the data-based diagnosis and support system for at-risk students.

The remainder of the paper is as follows. Section 2 describes relates approach to predict university dropout. Sections 3 and 4 respectively describe the methodology followed in our study and the results obtained. Finally, Section 5 presents the conclusions and suggests some future work guidelines.

2. Literature Review

The first student dropout-retention models were developed during the 70s. They followed different approaches, such as: psychological, sociological, interactionist, organizational and economic. Personal, family, institutional and academic variables are usually considered in these studies (Donoso & Schiefelbein, 2007; Fonseca & García, 2016; Suárez-Montes & Díaz-Subieta, 2015).

The *sociological approach* assesses the influence of external factors. Spady's model (Spady, 1970) considered that social integration in the university determines the student's commitment or their decision to drop out. Other factors considered are the influence of the family, expectations and demands that affect the potential, and academic performance of the student.

The *interactionist and organizational approaches* base student integration at university on their academic and social interaction. For example, Tinto's model (Tinto, 1975) conceived that the decision to drop out is affected by the social and academic interactions that students have during their higher education and the purposes, goals and commitment to the institution. This model considered that most of the dropout decisions are voluntary and are produced by an inadequate integration of the student who abandons both the social and intellectual environment of the institution.

Bean's model considered the following factors (Bean, 1980): (i) academic performance and integration, (ii) psycho-social (goals and interaction), and (iii) environmental (financing and opportunities). These factors influence performance, adaptation, and commitment to the institution. According to the positive or negative valence of these factors, students decide to stay or leave the institution.

In (Pascarella & Terenzini, 1980) the causes associated with dropout were classified into five groups of variables: (i) pre-university background of the students, (ii) structural and organizational characteristics of the institution, (iii) institutional, (iv) student interactions with socialization agents, and (v) quality of student effort.

The *psychological approach* is related to the students' characteristics and attributes. (Ethington, 1990) considered that the aspirations, values and expectations of students' success are determined by the stimulus and family support, the academic self-concept and level of difficulty that they present in the studies conditioned by the family background, and previous academic performance before entering university. The influence of personality traits has been also analyzed in recent studies (Alkan, 2014; Migali & Zucchelli, 2017). For instance, Migali and Zucchelli (2017) concluded that introversion, and to a lesser extent neuroticism, are individually associated with higher probabilities of dropping out.

There are also models focused on economic factors and cost-benefit ratio. For example, Cabrera, Nora, Terenzini, Pascarella and Hagedorn (1999) related students' dropout or persistence of the student in the university with their previous academic ability and the socioeconomic factors, the estimation of the costs and benefits of studying a career, the experience and academic performance acquired, as well as the possibility of financial support.

These initial models defined the basis for subsequent research and do not present contradictory proposals, but rather feed on each other and it is possible to combine them in different ways. In addition, there is a variable common to all of them, related to the academic ability of the student prior to their university admission.

In the 90s, countries such as Argentina, Brazil, Colombia, Cuba, Chile, Costa Rica and Mexico began the implementation of national evaluation and accreditation systems, including within their quality indicators, academic efficiency and retention of the students (Costa et al., 2018; Villanueva, Bentancur, Lacerda & González, 2008). These evaluations identified the need to continue the studies to understand and address the problem of desertion. The stage from 1970 to 2000 can be considered as the stage of proposal and consolidation of the initial models of student dropout/retention.

More recent proposals have a focus on prediction. In a review of his model, Tinto (2010) recognized four institutional conditions associated with student retention: expectations, support, feedback, and participation. His study deals with how Higher Education Institutions can contribute to the academic and/or social integration of the student, emphasizing the importance of developing expectations for success through academic, social and financial counseling, and support. It also emphasizes the need to use several evaluation methods and immediate and systematic feedback to the student, which guarantee an “early warning” to provide the necessary support. Also, it refers to the participation and commitment of the student with their learning.

On the other hand, Kerby (2015) proposed a predictive model of desertion applying classical sociological theory. The study incorporated Spady, Tinto and Bean models, internal factors (culture and institutional climate) and external factors (national and educational climate). In this sense, Spady, Tinto and Bean models use academic, personal, family and institutional factors, identifying them as pre-university (referred to students) and internal factors (referred to the institution) affect the adaptation of the student. The interaction between these factors allows internal factors to adapt to the needs of students, as external factors change.

In relation to the organizational approach, Fonseca & García (2016) analyzed Tinto, Bean, Pascarella and other models. They also provided the characteristics of quantitative studies of these authors using a large number of samples, both from students and institutions, using statistical correlation techniques, logistic regression and factor analysis; as well as a study of criticisms of these studies either to improve them or to make new proposals.

In spite of the different criteria, the first years of the degree are those of greater desertion. The causes related to academic performance, motivation for study and academic and social integration, caused by personal, family, socioeconomic or institutional factors are analyzed with greater emphasis. However, pedagogical or course-related variables have been less investigated. In our research we consider academic achievement in the most relevant subjects of the first year (specifically those related to Mathematics and Programming).

Several studies have examined students' dropout as a public health perspective linking education and health by examining risk and protective factors that might alter the relation between dropping out and subsequent negative outcomes (e.g., more criminal activity, poorer health, and lower tax contributions) (Lansford et al., 2016; Mussida, Sciulli & Signorelli, 2018; Ramsdal et al., 2013; Robison, Jagers, Rhodes, Blackmon & Church, 2017).

2.1. Use of Data Analytics

A widespread approach to study the causes of student dropout is to use educational data analytics to find the best performing features for classification between dropout and promotion (Araque, Roldán & Salguero, 2009; Chung & Lee, 2019). Even the authors that provided the initial models of desertion, which we surveyed in previous work (Lázaro, Callejas & Griol, 2017), used data mining to corroborate their results (Bean, 1980; Spady, 1970).

Recently, (Ullah, Alam, Mahiuddin & Rahman, 2019) used Naïve Bayes, Random Forest, and Logistic Regression to find a relationship between student desertion and student dissatisfaction with some university services. Other authors use data mining to predict student performance during the end-of-semester exams using their results in systematic educational activities such as class exams, seminars, homework, or laboratory work (Baradwaj & Pal, 2011). Chies, Graziosi and Pauli (2014) analyzed the factors influencing dropout, defined as not enrolling in the second year of the three years bachelor programs at the University of Trieste. The random lasso procedure was used to model the probability of dropout taking account of individual characteristics, university performance and job placement.

Regarding retention, different analytic methods have been used to predict whether students will stay through their university degree. For example, (Nandeshwar, Menzies & Nelson, 2011) found family background, family's socioeconomic status and the results of the exams to be determinant.

The need for graduates in Science, Technology, Engineering and Mathematics (STEM) is constantly growing. However their related studies experiment an increased student dropout, which has been recently reported and analyzed.

For instance, Villwock, Appio and Andreta (2015) have investigated the causes of desertion in the degree of Mathematics, using socioeconomic factors and the result in the courses taken by the students, obtaining that in the first year the most determinant subjects are Differential and Integral Calculus.

In the particular case of Computer Science, Costa, Fonseca, Santana and de Araújo (2017) make a comparative study of the permanence of students in distance or face-to-face courses on campus, using different student data, such as: age, sex, marital status, city, income, student record, period, class, semester, campus, year of enrollment in the course, state of the discipline and academic performance, attendance to on-site classes and in the case of distance learning, the frequency of accesses in the system.

Badr, Algobail, Almutairi and Almutery (2016) also performed a study to predict academic performance in a Programming course, using only as predictor variables performance in Mathematics and English courses, concluding that the performance in English courses has a greater predictive effect on performance in Programming.

In addition, Lacave, Molina and Cruz-Lemus (2018) have very recently completed a study about the causes of abandonment in the Computer Science degree at the University of Castilla-La Mancha in Spain, which is close to 40%. The qualification obtained in the university entrance examination was considered a predictive factor of abandonment when only their age was analyzed; when they had a scholarship, the province and the option in which they selected the studies were the most relevant. The academic performance in the subjects studied was also one of the most predictive variables.

Other authors analyze dropout using only data available when students enroll at university. For instance, in the context of technical studies, Nagy and Molontay (2018) used 15,825 cases and obtained a high classification accuracy (79%) with features related to the study program, whether the student was enrolled for the first year and their financial situation. They also obtained that successful students were not only the good at mathematics and science at high school, but those who also had a good performance in humanities.

Other studies have proposed economic and financial engineering methods to predict students' dropout. Barra and Zotti (2017) propose a stochastic approach to estimate the efficiency of a public university in Italy. To do this, they used data on the students' socioeconomic and educational background, such as the type of secondary school they attend, the gender, age, financial conditions of the families, and the distance to the university campus. The authors concluded that students with better socioeconomic and educational backgrounds are better integrated into the university system at said University. However, taking into account these requirements in admission policies would be in contradiction with resolution 70/1 of the United Nations General Assembly "Transforming our world: the 2030 Agenda for

Sustainable Development”, in which the objective 4 is aimed at guaranteeing an inclusive and equitable education for all: “By 2030, ensure equal access for all women and men to affordable and technical quality, vocational and tertiary education, including university” (United Nations, 2015).

In the Cuban context the expectations regarding retention results in Computer Engineering studies are not met. However, this phenomenon is not sufficiently addressed in the literature. We performed a search in Web of Science and Scopus with the criterion: “data* mining*, dropout and university” and obtained 40 papers in Web of Science (13 related to Engineering, from which 9 specifically addressed dropout in Computer Science) and 52 in Scopus (12 related to Engineering and 4 to Computer Science). None of them was focused on the Cuban context.

A search in Google Academic only retrieved 3 papers by Cuban authors, 2 in the area of Medicine (López, Marín & García, 2012; Pernas-Gómez, Sierra-Figueroa, Fernández-Sacasas, Miralles-Aguilera & Diego-Cobelo, 2009) and a comprehensive study that reported more than 4,000 dropouts in 17 higher education institutions in (Delgado, María & Quijada-González, 2012). The authors highlighted that “the increment of retention rates [in the country] is an unsolved problem which complexity requires a comprehensive treatment through rigorous research”.

Thus, we have performed this study to find avenues for the early prediction of dropout in the Computer Engineering degree. The general aim is providing Cuban lecturers and administrators the arguments necessary to transform the factors that provoke student dropout.

3. Method

We gathered institutional information from 456 students from all Cuban provinces enrolled in Computer Engineering studies (*Ingeniería Informática*) in the academic year 2013-2014. From them, 279 promoted to the second year, 83 repeated the first year and 94 dropped out. In this cohort, 47.42% of the students graduated in their corresponding year (course 17/18), 16.29% will probably graduate during the present year (course 18/19) and 36.29% have dropped out during the 5 years of the degree. From non-promoting students, 56.96% dropped out in the first year.

Our study comprises two sets of experiments: considering two classes (promoted or not promoted) and three classes (promotion, repetition, dropout). In Cuban Higher Education, a student that fails two subjects in the same semester or two or more in the whole academic year, is assessed and considered to repeat the year. Each student has the opportunity to repeat up to two years during the study plan and can only repeat the same level once.

With respect to the first research question, we aim to detect the maximum number of students in risk of dropout or repetition (not promoting). Our goal is not only to achieve the best classification performance possible, the best results for the study would be those with a higher recall in the dropout class, i.e. the maximum number of students in risk of dropout are identified; even if the precision is not as high (even if we predict as dropout students that finally promote). Similarly, it is desirable to attain a high precision of the promotions (not considering erroneously that a student in risk of dropout is going to promote), even if their recall is not as high, i.e. even if we do not identify as such all the promoted students. The overall idea is then to offer help to the higher number of students in risk as possible, even if we “unnecessarily” offer help to students that are finally going to promote.

Regarding the second research question, our purpose is to detect dropout risk as soon as possible. In the literature surveyed we have not found previous research that compares results at different moments during the first year of studies. That is why we propose to consider three types of analyses that comprise information gathered at different times: at registration time, after the first semester and after the first year. Each analysis comprises a series of features described in detail in the following section.

With respect to the machine learning approaches used, the most common in the previously cited literature are neural networks, decision trees, Bayesian approach and logistic regression, and the most widespread

software tools used are WEKA and R (Costa et al., 2017; Miranda & Guzmán, 2017; Nandeshwar et al., 2011). For example, very recently Vila, Cisneros, Granda and Ortega (2019) used them to detect dropout patterns in Ecuador, Ullah et al. (2019) used them to study aspects of student dissatisfaction in Bangladesh, and Mohamed and Waguih (2018) used them to propose a counseling model for students to select academic degrees. Furthermore, decision trees allow visualizing the classification process and identifying relevant features for decision-making.

Thus, for each experiment set we employed a J48 decision tree and a multilayer perceptron (MLP). For all experiments we used a 5-fold cross-validation approach, so all pieces of data are adequately used for training and test. As in previous studies - e.g. (Villwock et al., 2015), we used WEKA to run the experiments (Amaya, Barrientos & Heredia, 2015; Mohamed & Waguih, 2018; Vila et al., 2019).

The features proposed to predict dropout have been arranged in 3 groups (see Table 1): pre-registration, first semester and first year features. Each group comprises the previous ones, i.e. first year features comprise the pre-registration and first semester features and adds new features related to the second semester; while the first semester features comprise the admission features and features related to the first semester.

Group	Features	Type	Code
Pre-registration	Gender	Dichotomous	Gender
	Province	Nominal	Province
	Entry source	Dichotomous	EntryS
	Previous academic index	Interval	PrevAcInd
	Mark in <i>Mathematics</i> in the admission examination	Interval	MathsEntry
	Degree option rank	Ordinal	OptionRank
First semester	All pre-registration features	-	-
	Mark of the subject <i>Discrete Mathematics I</i>	Interval	DiscMathI
	Mark of the subject <i>Mathematics I</i>	Interval	MathI
	Mark of the subject <i>Lineal Algebra</i>	Interval	LinAlg
	Mark of the sub. <i>Introduction to Computer Science</i>	Interval	IntroCS
	Mark of the subject <i>Introduction to Programming</i>	Interval	IntroProg
	% Mathematics subjects passed in the 1st semester	Ratio	PerMathSem1
	% Programming subjects passed in the 1st semester	Ratio	PerProgSem1
	% subjects passed in total in the 1st semester	Ratio	PerPassedSem1
First year	All previous features	-	-
	Mark of the subject <i>Discrete Mathematics II</i>	Interval	DiscMathII
	Mark of the subject <i>Mathematics II</i>	Interval	MathII
	Mark of the subject <i>Programming I</i>	Interval	ProgI
	% Mathematics subjects passed in the 2nd semester	Ratio	PerMathSem2
	% Programming subjects passed in the 2nd semester	Ratio	PerProgSem2
	% subjects passed in total in the 2nd semester	Ratio	PerPassedSem2
	% Mathematics subjects passed in the year	Ratio	PerMathTotal
	% Programming subjects passed in the year	Ratio	PerProgTotal
	% subjects passed in total in the year	Ratio	PerPassedTotal

Table 1. Features considered for the detection of dropout

The pre-registration features used are similar to previous studies (Barra & Zotti, 2017; Nagy & Molontay, 2018), but adapted to the Cuban context:

Gender: 0 male and 1 female.

Province of origin: Students from all Cuban provinces have access to the Computer Engineering degree. The 15 provinces of the country are: 1 Havana, 2 Pinar del Río, 3 Artemisa, 4 Mayabeque, 5 Matanzas, 6 Villa Clara, 7 Sancti Spíritus, 8 Cienfuegos, 9 Ciego de Avila, 10 Camagüey, 11 Las Tunas, 12 Holguín, 13 Granma, 14 Santiago de Cuba, and 15 Guantánamo.

Entry source: Pre-university (also called baccalaureate) is studied before entering university, this training is mostly taught in the Urban Pre-University Institutes (IPU) that exist in all the provinces of the country, but it is also taught in other types of centers with more specific training (e.g. sports, military training) with lower enrollment. This feature can take two values: 1 IPU and 2 others.

Academic index prior to admission: It is important to consider the results obtained by the students during the pre-university level. This index is calculated by averaging the result of all the subjects received in this stage. Its value is a number from 0 to 100 (the higher the better).

Mark obtained in the Mathematics examinations for university admission: The only requirements to enter universities in Cuba are to have completed pre-university education and to pass three entrance tests: Mathematics, Spanish and History of Cuba. These tests are the same for all students in the country and are performed simultaneously. The scores of these tests use the scale from 0 to 100 points, students pass the exams with grades greater or equal to 60 points. For our study we consider the grade obtained by the student in Mathematics, as this discipline has a great impact in the first two years of the Computer Engineering degree.

Degree option rank: Prior to entering Higher Education, students must choose 10 career options in order of preference. After completing the entrance exams, the definitive granting is made ordering the students according to the average of their grades, so according to these results, sometimes the students may be assigned the least desired career. Hence the importance of studying the incidence of the option in which the career was requested in the student's decision to remain or dropout during or at the end of their first year of studies.

Secondly, we have used academic performance features, which have been used as indicators in previous studies (Dužević, 2015). In our case, we have focused on Mathematics and Programming subjects and also differentiated first and second semester subjects to be able to discern whether it is possible to predict dropout before the second semester.

Academic performance: The academic performance in a subject in Cuban universities is evaluated using the following categories: excellent (5), good (4), fair (3) and poor (2). The ratings express different degrees of mastery of the objectives, where grade 2 indicates that the student does not reach the minimum level required. For our experiments we analyze the incidence of academic performance in the first-year subjects of Mathematics and Programming in student desertion.

In the first year, students receive 5 basic subjects of the Mathematics discipline: Discrete Mathematics I, Mathematics I and Linear Algebra in the first semester, Discrete Mathematics II and Mathematics II in the second semester. They also receive 3 subjects of the specialty: Introduction to Computer Science and Introduction to Programming in the first semester and Programming I in the second semester. For this study we have considered the grades obtained in each of these subjects and have also calculated the percentage of passed subjects in different moments and disciplines:

- Percentage passed in the first semester in Mathematics
- Percentage passed in the first semester in Programming
- Percentage passed in the first semester in total
- Percentage passed in the second semester in Mathematics
- Percentage passed in the second semester in Programming
- Percentage passed in the second semester in total
- Percentage passed in total Mathematics

- Percentage passed in total Programming
- Percentage passed in total (two semesters)

4. Results

As previously described, we have completed two groups of experiments to predict dropout: considering three classes (promotion, repetition and dropout) and two classes (promotion and not promotion). The following subsections show the results obtained.

4.1. Analysis Considering Three Classes: Promotion, Repetition and Dropout

Features	Method	Accuracy	Precision	Recall	F-Measure	Class
<i>Pre-registration</i>	J48	59.87%	0.377	0.309	0.339	Dropout
			0.375	0.253	0.302	Repetition
			0.690	0.799	0.741	Promotion
			0.568	0.599	0.578	Weighted Avg.
	MLP	60.53%	0.333	0.223	0.268	Dropout
			0.395	0.181	0.248	Repetition
			0.676	0.860	0.757	Promotion
			0.554	0.605	0.563	Weighted Avg.
<i>First semester</i>	J48	85.08%	0.742	0.702	0.721	Dropout
			0.701	0.566	0.627	Repetition
			0.917	0.986	0.950	Promotion
			0.841	0.851	0.844	Weighted Avg.
	MLP	84.86%	0.747	0.755	0.751	Dropout
			0.711	0.651	0.679	Repetition
			0.919	0.939	0.929	Promotion
			0.846	0.849	0.847	Weighted Avg.
<i>First year</i>	J48	89.91%	0.814	0.745	0.778	Dropout
			0.762	0.771	0.766	Repetition
			0.965	0.989	0.977	Promotion
			0.897	0.899	0.898	Weighted Avg.
	MLP	96.49%	0.815	0.798	0.806	Dropout
			0.771	0.771	0.771	Repetition
			0.972	0.978	0.975	Promotion
			0.903	0.904	0.903	Weighted Avg.

Table 2. Results considering three classes

The results for three classes are summarized in Table 2. As can be observed, when using only the data corresponding to pre-registration features, it is difficult to predict whether the students will dropout or repeat. The maximum accuracy is 60.53% and the maximum dropout recall is 0.31. The class with higher recall corresponds to the promoted students (0.79). However, its precision is relatively low (0.69), this indicates that promoting students are almost correctly classified (222 out of 279 are correctly classified), but a considerable number of non-promoting students are classified as promoting (100). This gives an idea that the classification learned is similar to a baseline that classifies all students as promoting, which would obtain an accuracy of 61.18%. This can be explained by the natural imbalance of the categories, as there will be always more individuals in the promoting category, and so a baseline that always categorizes a student as promoting already achieves a high accuracy.

Considering also the first semester features, the MLP obtains an accuracy of 84.87% and it is possible to predict 75.5% of the dropout cases and 65.1% of repeating students, with a precision of promoting

students of 0.919. Thus, compared to the scenario in which only the pre-registration features were used, there are less students erroneously classified as promoting (25/279) and more correctly classified as dropping out (66/94) or repeating (47/83). In addition, confusions for these two classes are mostly among them (classifying a repeating student as dropping out and vice versa) and not with the promoting class. Although accuracy is lower for J48, the recall of dropouts and repetitions (non-promoting students) is better than with the MLP.

When taking into account all features (corresponding to the whole year), the MLP method obtains the best results with 90.35% accuracy and the recall for dropout and repetition is 0.798 and 0.771 respectively.

The most convenient approach for the aim of this study would be to be able to predict non-promoting students before the end of the year (no longer than the first semester) in order to be able to prevent abandonment. In addition, although all classification errors had the same impact for the accuracy calculated in Table 2, not all have the same impact for decision makers, as classifying a student in risk of dropout as promoting is worse than classifying them as repeating, in terms of the personalized help that could be provided to them. This is why we performed a second group of experiments where the classes dropout and repetition were grouped into a single category of non-promoting students.

4.2. Analysis Considering Two Classes: Promoting, not Promoting

The results of the classification with two classes (promoting and not promoting) are shown in Table 3. As can be observed, using only pre-registration features, the best result is obtained with J48 with 68.86% accuracy, recall of non-promoting of 0.644 and precision of promoting of 0.760. Thus, considering only 2 classes, we are able to predict that a student is in risk of not promoting in 64.4% of the cases, versus 33% with the three-class classifier. As we are considering only their situation prior to the start of their studies, this classification approach would help to adopt early measures to avoid dropout.

Features	Method	Accuracy	Precision	Recall	F-Measure	Class
<i>Pre-registration</i>	J48	68.86%	0.591	0.644	0.616	Not promoting
			0.760	0.717	0.738	Promoting
			0.695	0.689	0.691	Weighted Avg.
	MLP	66.67%	0.573	0.554	0.563	Not promoting
			0.723	0.723	0.730	Promoting
			0.665	0.667	0.666	Weighted Avg.
<i>First semester</i>	J48	93.85%	0.975	0.864	0.916	Not promoting
			0.920	0.986	0.952	Promoting
			0.941	0.939	0.938	Weighted Avg.
	MLP	96.71%	0.881	0.876	0.878	Not promoting
			0.921	0.925	0.923	Promoting
			0.906	0.906	0.906	Weighted Avg.
<i>First year</i>	J48	96.71%	0.966	0.949	0.957	Not promoting
			0.968	0.978	0.973	Promoting
			0.967	0.967	0.967	Weighted Avg.
	MLP	96.49%	0.960	0.949	0.949	Not promoting
			0.968	0.975	0.971	Promoting
			0.965	0.965	0.965	Weighted Avg.

Table 3. Results considering two classes

Considering also the first semester data, although accuracy is higher with J48 (93.8%), for our aim the best alternative is obtained with the MLP as although accuracy is slightly lower (0.905), it predicts 87.6% of

non-promoting students, with a precision of 0.921 of the promoting students, parameters that are worse with J48.

When taking into account all features (also second semester), the best result is achieved with J48, with a total accuracy of 96.7% (much higher compared with 66.8% only with enrolment data and 93.85% with the data only up to the first semester). However, the attributes that determine these results are obtained after the end of the first year and thus do not allow an early intervention.

To obtain a closer perspective of the classification, we performed an analysis of the decision trees obtained for the two classes classification (promoting and not promoting).

Figure 1 shows the decision tree obtained with all features. As can be observed, it is possible to immediately classify 171 cases (96.6% of not promoting students) only with the feature that considers the average number of subjects that the student has passed. Unfortunately, this information is only available at the end of the academic year when it is too late to provide adequate prevention mechanisms.

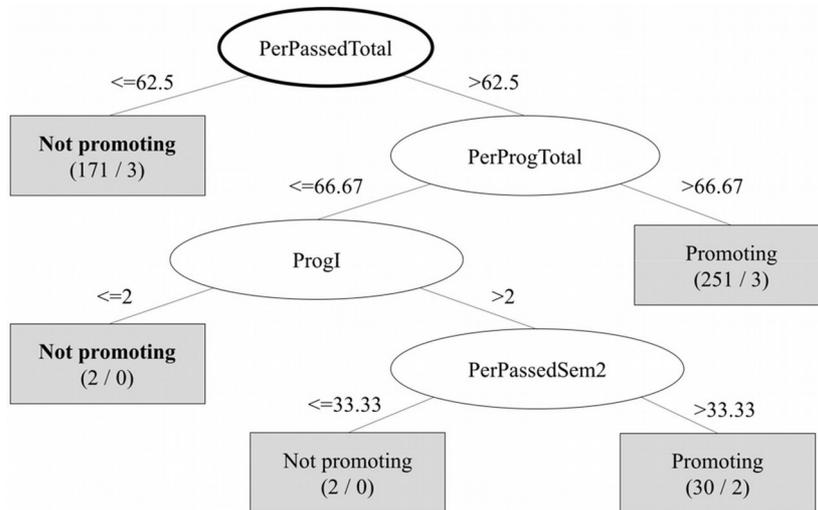


Figure 1. Decision tree corresponding to all features

When considering features only up to the first semester, the resulting tree is depicted in Figure 2. The result only considers the feature that corresponds to the average number of subjects passed in the first semester, which already classifies correctly 93.8% of the cases. Despite the high classification rate attained, this result is not helpful either for decision makers, as it has a clear correspondence with the promotion regulations of Cuban Higher Education explained in Section 3.

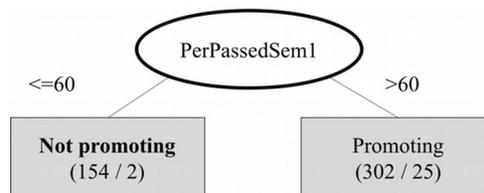


Figure 2. Decision tree corresponding to first semester features

Figure 3 shows the decision tree with only the features prior to enrolment. As can be observed, 50.3% of the cases are correctly classified considering only the features prior Mathematics academic index, province, career option and mark of examination in Mathematics for admission. Although most non-promoting students have a lower previous academic index (≤ 91.45) and are not from La Habana, the correct

classification of the instances entails a complex route in the tree considering a large number of variables, obtaining concretely a 68.86% accuracy.

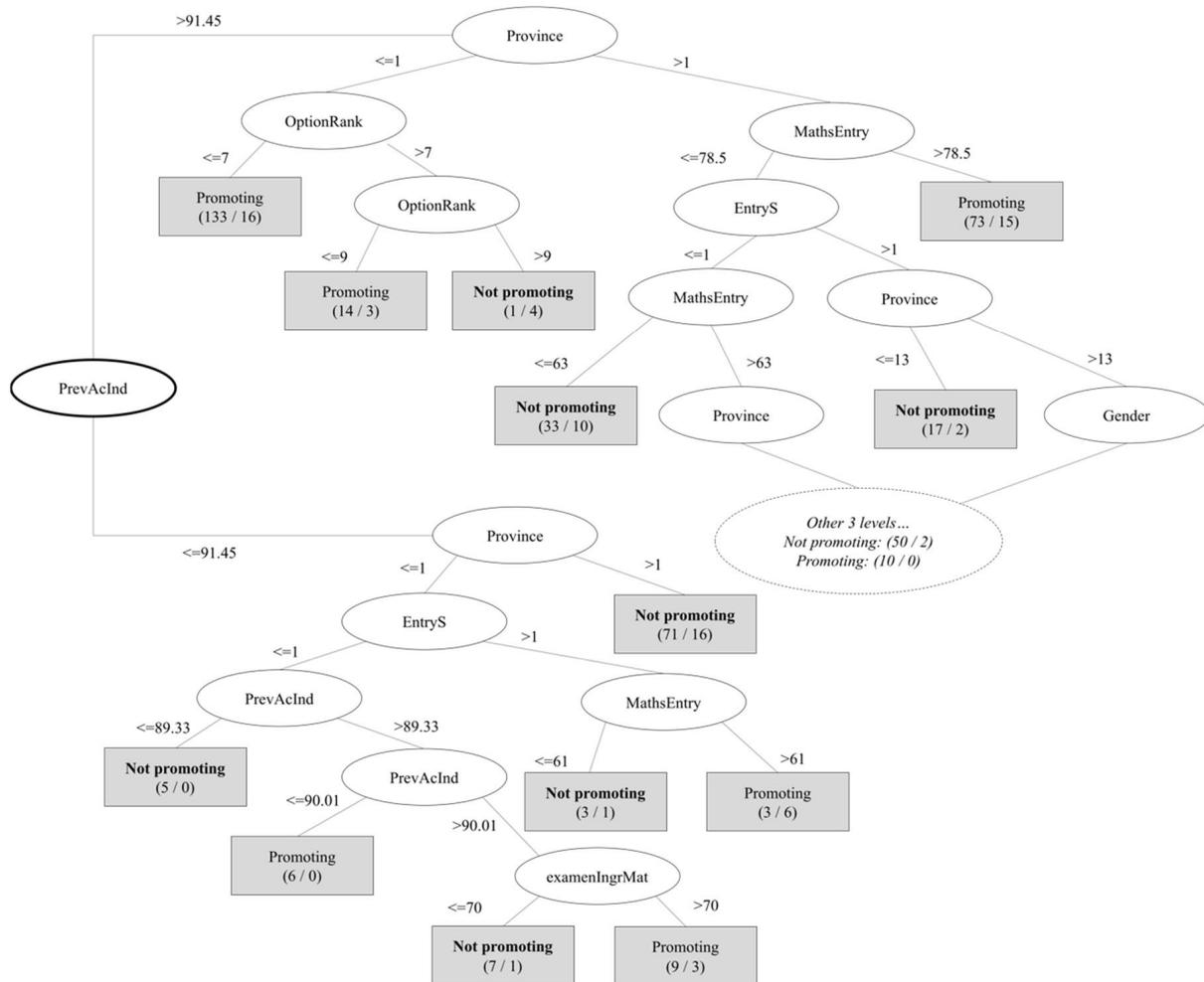


Figure 3. Decision tree corresponding to pre-registration features

Lacave et al. (2018) obtain similar results with a smaller sample to predict dropout in Computer Science within the Spanish context. In their study, the previous academic index also has a prominent role. With pre-registration features, Nagy and Molontay, (2018) obtained an accuracy of 63% using even more features (e.g. previous performance in Mathematics, Literature and foreign languages). With fewer features we obtain a 68.86% accuracy.

Considering the data of the full year, we obtain similar results to (Vila et al., 2019), a study in Ecuador with different features (age, average marks and disability information) that obtained 97% accuracy (our result is 96.71%).

5. Conclusions

Dropout is a very relevant challenge for Higher Education institutions with important implications in society. There is a long tradition of scientific research addressing the topic that has produced relevant models. However, practical results are in most cases not replicable between countries and institutions. Recently, machine-learning approaches have offered the possibility to process institutional data to identify the most relevant features that may allow the detection of students at risk.

We have presented an educational data analytics study about student dropout in Computer Engineering studies in Cuba, addressing two questions: i) First, whether it is possible to accurately predict which

students are in risk of dropout; and ii) whether such prediction can be performed early, before the end of the first academic year or even at the time of enrolment.

Our experimental results show that it is possible to determine whether a student is in risk of dropout after the first year of studies with 96.71% accuracy. When considering only pre-registration features, the accuracy is 68.86%, which enhances the results of the literature and shows it is possible to have information about the risk of dropout at the beginning of the first year. When first semester variables are considered, the accuracy rises to 96.71%, which is very convenient for teachers and policy makers to adopt early measures.

There exist multiple approaches in the literature related to institutional actions aimed at student retention. Tutoring and mentoring actions have proved to be successful retention strategies in different contexts (Alzen, Langdon & Otero, 2018; Araque et al., 2009; Elster, 2014; Graffigna, Hidalgo, Jofré, Berenguer, Moyano & Esteybar, 2014). Other authors propose the use of active teaching and learning methodologies (Canedo, Santos & Leite, 2018; Paimin, Prpic, Hadgraft & Alias, 2017; Zehetmeier, Axel, Bruggemann-Klein, Thurner & Schlierkamp, 2015). The role of technology and hands-on practices has also been highlighted in the literature (Kori, Pedaste, Nüitsoo, Kuusik, Altin, Tõnisson et al., 2015), as well as the importance of psychological support (Ahmed, Kloot & Collier-Reed, 2015; Teixeira de Mello, de Melo & de Mello Filho, 2015).

For future work we will explore these possibilities to design tutoring actions rooted on technology in the context of the Computer Engineering studies in Cuba. The dropout prediction approach presented will be implemented as a module embedded in the University Management System, allowing to identify students at risk of dropout and selecting the best suited tutoring actions and the actors that should participate in accordance with the predictive factors described.

With respect to the limitations of the study, we must consider that student desertion does not necessarily imply academic failure, as personal interests may guide some dropouts. Tinto (1982) already highlighted this issue and other authors consider different types of dropout, e.g. planned vs. derived by academic problems and lack of motivation (Zajac & Komendant-Brodowska, 2018). For future work we plan to conduct a qualitative study with interviews to students who dropped out in order to be able to identify the causes and distinguish between planned and undesired dropout.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- Ahmed, N., Kloot, B., & Collier-Reed, B.I. (2015). Why students leave engineering and built environment programmes when they are academically eligible to continue. *European Journal of Engineering Education*, 40(2), 128-144. <https://doi.org/10.1080/03043797.2014.928670>
- Alkan, N. (2014). Humor, Loneliness and Acceptance: Predictors of University Drop-out Intentions. *Procedia - Social and Behavioral Sciences*, 152, 1079-1086. <https://doi.org/10.1016/j.sbspro.2014.09.278>
- Allen, J., Robbins, S.B., Casillas, A., & Oh, I.S. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education*, 7(49), 647-664. <https://doi.org/10.1007/s11162-008-9098-3>

- Alzen, J.L., Langdon, L.S., & Otero, V.K. (2018). A logistic regression investigation of the relationship between the Learning Assistant model and failure rates in introductory STEM courses. *International Journal of STEM Education*, 5(1), 56. <https://doi.org/10.1186/s40594-018-0152-1>
- Amaya, Y., Barrientos, E., & Heredia, D. (2015). Student Dropout Predictive Model Using Data Mining Techniques. *IEEE Latin American Transactions*, 13(9), 3127-3134. <https://doi.org/10.1109/TLA.2015.7350068>
- Araque, F., Roldán, C., & Salguero, A. (2009). Factors influencing university drop out rates. *Computers & Education*, 53(3), 563-574. <https://doi.org/10.1016/j.compedu.2009.03.013>
- Badr, G., Algobail, A., Almutairi, H., & Almutery, M. (2016). Predicting Students' Performance in University Courses: A Case Study and Tool in KSU Mathematics Department. *Procedia - Procedia Computer Science*, 82(March), 80-89. <https://doi.org/10.1016/j.procs.2016.04.012>
- Baradwaj, B.K., & Pal, S. (2011). Mining Educational Data to Analyze Students' Performance. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2(6), 63-69. <https://doi.org/10.14569/IJACSA.2011.020609>
- Barra, C., & Zotti, R. (2017). What we can learn from the use of student data in efficiency analysis within the context of higher education? *Tertiary Education and Management*, 23(3), 276-303. <https://doi.org/10.1080/13583883.2017.1329450>
- Bean, J.P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12(2), 155-187.
- Braxton, J., & Hirschy, A.S. (2005). Theoretical Developments in College Student Departure. In *College Student Retention: Formula for Student Success* (67-87). Greenwood Press.
- Burgos, C., Campanario, M.L., Pena, D. de la, Lara, J.A., Lizcano, D., & Martinez, M.A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers and Electrical Engineering*, 66, 541-556. <https://doi.org/10.1016/j.compeleceng.2017.03.005>
- Cabrera, A.F., Nora, A., Terenzini, P.T., Pascarella, E., & Hagedorn, L.S. (1999). Campus Racial Climate and the Adjustment of Students to College: A Comparison between White Students and African-American Students. *The Journal of Higher Education*, 70(2), 134-160. <https://doi.org/10.1080/00221546.1999.11780759>
- Canedo, E.D., Santos, G.A., & Leite, L.L. (2018). An Assessment of the Teaching-Learning Methodologies Used in the Introductory Programming Courses at a Brazilian University. *Informatics in Education*, 17(1), 45-59. <https://doi.org/10.15388/infedu.2018.03>
- Chies, L., Graziosi, G., & Pauli, F. (2014). Job Opportunities and Academic Dropout: The Case of the University of Trieste. *Procedia Economics and Finance*, 17, 63-70. [https://doi.org/10.1016/S2212-5671\(14\)00879-X](https://doi.org/10.1016/S2212-5671(14)00879-X)
- Chung, J.Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, 96, 346-353. <https://doi.org/10.1016/j.childyouth.2018.11.030>
- Costa, F.J.D, Bispo, M.D.S., & Pereira, R.D.C.D.F. (2018). Dropout and retention of undergraduate students in management: A study at a Brazilian Federal University. *RAUSP Management Journal*, 53(1), 74-85. <https://doi.org/10.1016/j.rauspm.2017.12.007>

- Costa, E.B., Fonseca, B., Santana, M.A., & de Araújo, F.F. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73, 247-256. <https://doi.org/10.1016/j.chb.2017.01.047>
- Delgado, B., María, J., & Quijada-González, M. (2012). Los sistemas de acceso, normativas de permanencia, y estrategias de tutoría y retención de estudiantes de educación superior en Cuba. *Informe Nacional*. Proyecto ACCEDES.
- Donoso, S., & Schiefelbein, E. (2007). Análisis de los modelos explicativos de retención de estudiantes en la Universidad: Una visión desde la desigualdad social. *Estudios Pedagógicos*, XXXIII(1), 7-27. <https://doi.org/10.4067/S0718-07052007000100001>
- Dužević, I. (2015). A conceptual framework for analysing the impact of influences on student engagement and learning. *Tertiary Education and Management*, 21(1), 66-79. <https://doi.org/10.1080/13583883.2014.1000368>
- Elster, D. (2014). First-Year Students' Priorities and Choices in STEM Studies—IRIS Findings from Germany and Austria. *Science Education International*, 25(1), 52-59.
- Ethington, C.A. (1990). A psychological model of student persistence. *Research in Higher Education*, 31(3), 266-269. <https://doi.org/10.1007/BF00992313>
- Fonseca, G., & García, F. (2016). Permanencia y abandono de estudios en estudiantes universitarios: Un análisis desde la teoría organizacional. *Revista de la Educación Superior*, 45(179), 25-39. <https://doi.org/10.1016/j.resu.2016.06.004>
- Freeman, T.M., Anderman, L.H., & Jensen, J.M. (2007). Sense of belonging in college freshmen at the classroom and campus levels. *The Journal of Experimental Education*, 3(75), 203-220. <https://doi.org/10.3200/JEXE.75.3.203-220>
- Graffigna, A.M., Hidalgo, L., Jofré, A., Berenguer, M.D.C., Moyano, A., & Esteybar, I. (2014). Tutorial Practice as a Strategy of Retention at the School of Engineering. *Procedia - Social and Behavioral Sciences*, 116, 2489-2493. <https://doi.org/10.1016/j.sbspro.2014.01.598>
- Gregori, P., Martínez, V., & Moyano-Fernández, J.J. (2018). Basic actions to reduce dropout rates in distance learning. *Evaluation and Program Planning*, 66, 48-52. <https://doi.org/10.1016/j.evalprogplan.2017.10.004>
- Hausmann, L.R., Schofield, J.W., & Woods, R.L. (2008). Sense of belonging as a predictor of intentions to persist among African American and White first-year college students. *Research in Higher Education*, 7(48), 803-839. <https://doi.org/10.1007/s11162-007-9052-9>
- IESALC-UNESCO (2007). *Informe sobre la Educación Superior en América Latina y el Caribe 2000-2005*. Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC). Available at: http://www.oci.es/salactsi/informe_educacion_superiorAL2007.pdf
- Itzhaki, Y., Itzhaky, H., & Yablon, Y.B. (2018). The contribution of parental and societal conditional regard to adjustment of high school dropouts. *Journal of Adolescence*, 62, 151-161. <https://doi.org/10.1016/j.adolescence.2017.11.014>
- Kerby, M.B. (2015). Toward a New Predictive Model of Student Retention in Higher Education. *Journal of College Student Retention: Research, Theory & Practice*, 17(2), 138-161. <https://doi.org/10.1177/1521025115578229>
- Kori, K., Pedaste, M., Nütsoo, M., Kuusik, R., Altin, H., Tõnisson, E. et al. (2015). Why do Students Choose to Study Information and Communications Technology? *Procedia - Social and Behavioral Sciences*, 191, 2867-2872. <https://doi.org/10.1016/j.sbspro.2015.04.249>

- Lacave, C., Molina, A.I., & Cruz-Lemus, J.A. (2018). Learning Analytics to identify dropout factors of Computer Science studies through Bayesian networks. *Behaviour & Information Technology*, 37(10-11), 993-1007. <https://doi.org/10.1080/0144929X.2018.1485053>
- Lansford, J.E., Dodge, K.A., Pettit, G.S., & Bates, J.E. (2016). A Public Health Perspective on School Dropout and Adult Outcomes: A Prospective Study of Risk and Protective Factors From Age 5 to 27 Years. *Journal of Adolescent Health*, 58, 652-658. <https://doi.org/10.1016/j.jadohealth.2016.01.014>
- Lázaro, N., Callejas, Z., & Griol, D. (2017). La deserción estudiantil en educación superior: S.O.S. en carreras de Ingeniería Informática. *VII CLABES*. Available at: <http://revistas.utp.ac.pa/index.php/clabes/article/view/1674>
- López, I., Marín, G., & García, M.E. (2012). Deserción escolar en el primer año de la carrera de Medicina. *Revista de Educación Médica Superior*, 26(1), 45-52.
- López, S., Carpeño, A., Arriaga, J., & Ruiz, M. (2016). Experiencias para el Fomento de las Vocaciones Tecnológicas entre Estudiantes de Enseñanza Secundaria. In *TICAI 2016: TICs para el Aprendizaje de la Ingeniería* (89-96). IEEE.
- Lotkowski, V.A., Robbins, S.B., & Noeth, R.J. (2004). The Role of Academic and Non-Academic Factors in Improving College Retention. *ACT Policy Report. American College Testing ACT Inc*, 7(48), 803-839. <https://doi.org/10.1037/e420492008-001>
- Migali, G., & Zucchelli, E. (2017). Personality traits, forgone health care and high school dropout: Evidence from US adolescents. *Journal of Economic Psychology*, 62, 98-119. <https://doi.org/10.1016/j.joep.2017.06.007>
- Miranda, M.A., & Guzmán, J. (2017). Análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos. *Formacion Universitaria*, 10(3), 61-68. <https://doi.org/10.4067/S0718-50062017000300007>
- Mohamed, M.H., & Waguih, H.M. (2018). A proposed academic advisor model based on data mining classification techniques. *International Journal of Advanced Computer Research*, 8(36), 129-136. <https://doi.org/10.19101/IJACR.2018.836003>
- Munizaga-Mellado, F R., Cifuentes-Orellana, M.B., & Beltrán-Gabrie, A.J. (2018). Retención y Abandono Estudiantil en la Educación Superior Universitaria en América Latina y el Caribe: Una Revisión Sistemática. *Education Policy Analysis Archives*, 26, 61. <https://doi.org/10.14507/epaa.26.3348>
- Mussida, C., Sciulli, D., & Signorelli, M. (2018). Secondary school dropout and work outcomes in ten developing countries. *Journal of Policy Modeling*. <https://doi.org/10.1016/j.jpolmod.2018.06.005>
- Nagy, M., & Molontay, R. (2018). Predicting Dropout in Higher Education based on Secondary School Performance. *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, 389-394. <https://doi.org/10.1109/INES.2018.8523888>
- Nandeshwar, A., Menzies, T., & Nelson, A. (2011). Learning patterns of university student retention. *Expert Systems with Applications*, 38(12), 14984-14996. <https://doi.org/10.1016/j.eswa.2011.05.048>
- Paimin, A.N., Prpic, J.K., Hadgraft, R.G., & Alias, M. (2017). Understanding student's learning experiences in higher education. *Proceedings of INTED2017 Conference*, 6670-6676. <https://doi.org/10.21125/inted.2017.1543>
- Pascarella, E.T., & Terenzini, P.T. (1980). Predicting Freshman Persistence and Voluntary Dropout Decisions from a Theoretical Model. *The Journal of Higher Education*, 51(1), 60-75. <https://doi.org/10.2307/1981125>

- Peña-Calvo, J.V., Inda-Caro, M., Rodríguez-Menéndez, C., & Fernández-García, C. (2016). Perceived Supports and Barriers for Career Development for Second-Year STEM Students. *Journal of Engineering Education*, 105(2), 341-365. <https://doi.org/10.1002/jee.20115>
- Pernas-Gómez, M., Sierra-Figueroa, S., Fernández-Sacasas, J.A., Miralles-Aguilera, E., & Diego-Cobelo, J.M. (2009). Principios estratégicos de la educación en Ciencias de la salud en Cuba (II): La pertinencia. *Revista Cubana de Educación Médica Superior*, 23(2), 1-12.
- Pritchard, M.E., & Wilson, G.S. (2003). Using emotional and social factors to predict student success. *Journal of College Student Development*, 1(44), 18-28. <https://doi.org/10.1353/csd.2003.0008>
- Ramsdal, G., Gurgens, R., & Wynn, R. (2013). Dropout and early unemployment. *International Journal of Educational Research*, 62, 75-86. <https://doi.org/10.1016/j.ijer.2013.06.011>
- Reason, R.D. (2003). Student variables that predict retention: Recent research and new developments. *Naspa Journal*, 4(40), 598-616. <https://doi.org/10.2202/0027-6014.1286>
- Robbins, S.B., Allen, J., Casillas, A., Peterson, C.H., & Lee, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of College Student Development*, 3(98), 598-616. <https://doi.org/10.1037/0022-0663.98.3.598>
- Robison, S., Jagers, J., Rhodes, J., Blackmon, B.J., & Church, W. (2017). Correlates of educational success: Predictors of school dropout and graduation for urban students in the Deep South. *Children and Youth Services Review*, 73, 37-46. <https://doi.org/10.1016/j.childyouth.2016.11.031>
- Rodríguez-Gómez, D., Feixas, M., Gairín, J., & Muñoz, J.L. (2012). Understanding Catalan University Dropout from a Comparative Approach. *Procedia - Social and Behavioral Sciences*, 46, 1424-1429. <https://doi.org/10.1016/j.sbspro.2012.05.314>
- Spady, W.G. (1970). Dropouts from Higher Education: An Interdisciplinary Review and Synthesis. *Interchange*, 1(1), 64-85. <https://doi.org/10.1007/BF02214313>
- Suárez-Montes, N., & Díaz-Subieta, L.B. (2015). Estrés académico, deserción y estrategias de retención de estudiantes en la educación superior. *Revista de Salud Pública*, 17(2), 300-313. <https://doi.org/10.15446/rsap.v17n2.52891>
- Teixeira de Mello, S.P., de Melo, P.A., & de Mello Filho, R.T. (2015). Study of the Dropout in Technological Education in a Higher Education Institution in Southern Brazil. *Eccos-Revista Científica*, 37, 181-196. <https://doi.org/10.5585/eccos.n37.4890>
- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research Winter*, 45(1), 89-125. <https://doi.org/10.3102/00346543045001089>
- Tinto, V. (1982). Defining dropout: A matter of perspective. *New Directions for Institutional Research*, 1982(36), 3-15. <https://doi.org/10.1002/ir.37019823603>
- Tinto, V. (2010). From theory to action: Exploring the institutional conditions for student retention. *Higher education: Handbook of theory and research*. https://doi.org/10.1007/978-90-481-8598-6_2
- Ullah, M.A., Alam, M.M., Mahiuddin, Md., & Rahman, M.M. (2019). Predicting Factors of Students Dissatisfaction for Retention. In Abraham, A., Dutta, P., Mandal, J. K., Bhattacharya, A., & Dutta, S. (Eds.), *Advances in Intelligent Systems and Computing Proceedings of IEMIS 2018* (755, 501-510). Springer. https://doi.org/10.1007/978-981-13-1951-8_45
- United Nations (2015). *Transforming our world: The 2030 Agenda for Sustainable Development (A/RES/70/1)*. United Nations. Available at: <https://sustainabledevelopment.un.org/post2015/transformingourworld>

- Vila, D., Cisneros, S., Granda, P., & Ortega, C. (2019). Detection of Desertion Patterns in University Students Using Data Mining Techniques: A Case Study. In Botto-Tobar, M., Pizarro, G., Zúñiga-Prieto, M., D'Armas, M., & Sánchez, M.Z. (Eds.), *Technology Trends. CITT 2018. Communications in Computer and Information Science* (895, 420-429). Springer, Cham. https://doi.org/10.1007/978-3-030-05532-5_31
- Villanueva, E., Bentancur, N., Lacerda, M. de, & González, M.D. (2008). Reformas de la educación superior: 25 propuestas para la educación superior en América Latina y el Caribe. *IESALC-UNESCO*, 241-295.
- Villwock, R., Appio, A., & Andreta, A.A. (2015). Educational Data Mining with Focus on Dropout Rates. *IJCSNS International Journal of Computer Science and Network Security*, 15(3), 17-23.
- Williams, B., Wankat, P.C., & Neto, P. (2018). Not so global: A bibliometric look at engineering education research. *European Journal of Engineering Education*, 43(2), 190-200. <https://doi.org/10.1080/03043797.2016.1153043>
- Zajac, T.Z., & Komendant-Brodowska, A. (2018). Premeditated, dismissed and disenchanted: Higher education dropouts in Poland. *Tertiary Education and Management*, 25(1), 1-16. <https://doi.org/10.1007/s11233-018-09010-z>
- Zehetmeier, D., Axel, B., Bruggemann-Klein, A., Thurner, V., & Schlierkamp, K. (2015). Finding Competence Characteristics among First Semester Students in Computer Science. *45th Annual Frontiers in Education Conference*, 1079-1087. <https://doi.org/10.1109/FIE.2015.7344201>

Published by OmniaScience (www.omniascience.com)

Journal of Technology and Science Education, 2020 (www.jotse.org)



Article's contents are provided on an Attribution-Non Commercial 4.0 Creative commons International License.

Readers are allowed to copy, distribute and communicate article's contents, provided the author's and JOTSE journal's names are included. It must not be used for commercial purposes. To see the complete licence contents, please visit <https://creativecommons.org/licenses/by-nc/4.0/>.