

# Prediction of university dropout through technological factors: a case study in Ecuador

## Predicción de la deserción estudiantil universitaria a través de factores tecnológicos: un caso de estudio en Ecuador

Mayra Susana ALBAN Taibe [1](#); David MAURICIO Sánchez [2](#)

Received: 27/06/2018 • Approved: 10/09/2018 • Published 28/12/2018

### Contents

[1. Introduction](#)

[2. Methodology](#)

[3. Results](#)

[4. Conclusions](#)

[Bibliographic references](#)

#### ABSTRACT:

Predicting dropout in universities has become a concern in several countries around the world. With the introduction of new information and communication technologies, new factors have appeared that influence student dropout in universities. This article proposes an approach to machine learning based on logistic regression techniques and decision trees and factors such as Internet addiction, addiction to social networks and addiction to technology, that affect the desertion of students in universities. As a result, it was obtained that the technique with the highest percentage of dropout precision was decision trees with 91.70%.

**Keywords:** Prediction of dropout, technological factors, machine learning

#### RESUMEN:

Predecir la deserción en las universidades se ha convertido en una preocupación en varios países del mundo. Con la introducción de las nuevas tecnologías de información y comunicación han aparecido nuevos factores que influyen en la deserción estudiantil en las universidades. Este artículo propone un enfoque de aprendizaje automático basado en las técnicas de regresión logística y árboles de decisión y en los factores adicción al internet, adicción a las redes sociales y adicción a la tecnología que afectan a la deserción de los estudiantes en las universidades. Como resultado se obtiene que la técnica que presenta mayor porcentaje de precisión de la deserción fue árboles de decisión con un 91.70%.

**Palabras clave:** Predicción de la deserción, factores tecnológicos, aprendizaje automático

## 1. Introduction

University desertion is considered a problem that affects higher education institutions around the world. It is a topic that generates controversy in the education context, where different actors such as managers, teachers and students are involved (Díaz Peralta, 2008).

Nowadays, the desertion rates are analyzed as quality criteria in the evaluation process and university accreditation, which in several cases imply academic and social changes (Djulovic

& Li, 2013). Student desertion research is motivated by its higher rates and economic costs that affect students and higher education institutions. In Chile 3 of 10 students drop out college before or at the end of the second semester (Argote & Jimenez, 2016), whilst in the United States approximately 50% of the students that start an undergraduate program are not able to finish their studies and obtain a professional degree. This problem is more serious in public universities than in private ones.

On the other side, the introduction of technology in the higher education entails the transformation of the educational process (Bustos Andreu & Nussbaum, 2009). According to MORENO and MOLINA (2014), the projects that define the ideological conceptions of education are influenced by science and technology. Moreover, the knowledge society at the universities is linked with the development of technological competencies that are a requirement in the present globalized world. According to Henríquez and Escobar (2016), the current education is influenced by the innovation of the information and communication technologies. These technologies can be utilized as a pedagogical mediation that allows students to find different ways on how to deal with university desertion Méndez-Estrada and Barrientos Llosa (2013).

It is important to emphasize that technological advances incorporate other factors that affect the students' decision on dropping out from college. Technological resources such as internet and social media when used inadequately can provoke negative effects to the student, triggering university desertion. For Blau (2011) and (Ghamari, Mohammadbeigi, Mohammadsalehi, & Hashiani, 2011), the misuse of the internet generates negative implications in the academic performance. Students are adopting digital technologies at early ages which make them sensitive to digital disturbances (Gencer & Koc, 2012).

As university students have more accessibility to internet usage, their vulnerability to abuse of information and communication technologies increases. Students have more freedom, longer periods of non-structured time and stimulus given by the educational environment to use internet which leads into the loss of education productivity.

Furthermore, Kittinger, Correia, and Irons (2012) indicate that social media is one of the most popular activities related with internet usage specially in university students. The misuse of social media is characterized as an addictive behavior within young people. Carbonell et al. (2012) study highlights the relation between psychological anguish and the inadequate internet usage and mobile phones which lead to depression, anxiety, insomnia. These conditions can mislead students from their education objectives.

Despite the existence of numerous researches related to the objectives of this work. The majority of studies aim to highlight the positive benefits of information and communication technology usage in the teaching-learning university process and leave behind the analysis of the negative influence that can cause the misuse of these technologies in the decision of university desertion.

Given the necessity of reducing student university drop outs, the following article's objective is to analyze the influence of factors related with the inadequate usage of internet, social media and technology in the university desertion and evaluate through data mining techniques if these factors can predict college drop out.

This article is comprised of four sections; the second section presents the methodology for the development of the research. The third section presents the results of the experimental process and finally the section four discusses the results and presents conclusions.

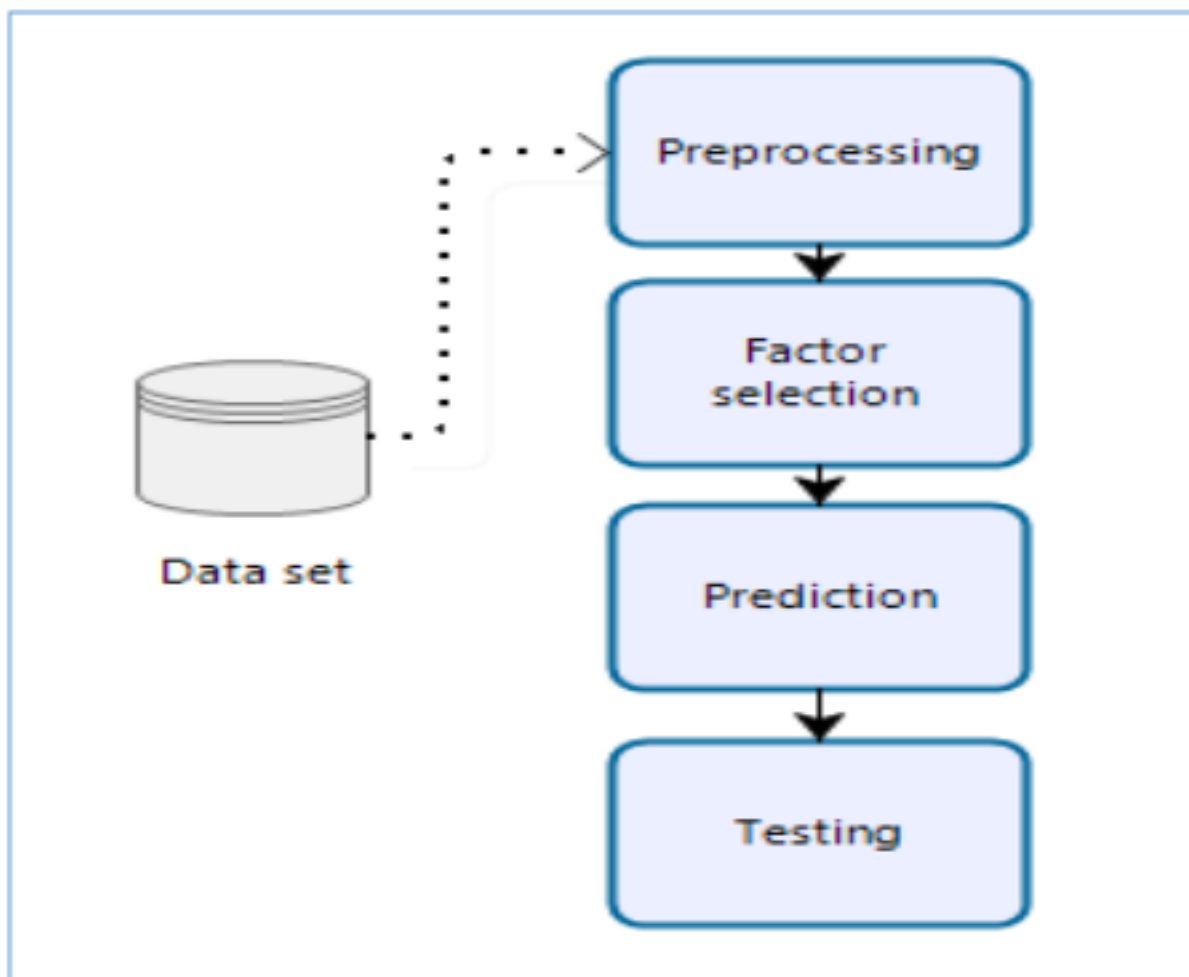
---

## 2. Methodology

The objective of the following investigation was to predict university student desertion in engineering students at public universities in Ecuador, using technological factors and the application of automatic learning based on logistic regression techniques and decision trees. For this reason, four phases of data mining were established, starting with preprocessing data until the testing phase, which are presented in Figure 1.

**Figure 1**

Prediction process of university student dropout



## 2.1. Dataset

A survey using Google forms was administered to 1178 students from computer science, electro mechanics, industrial and electric undergraduate careers. Taking into consideration the objectives of the research and the target population, the type of questions and their basic characteristics were determined because they were considered as variables or indicators of the information acquired (Murillo, 2006).

The survey was divided into three sections; the first section is related with the characteristics of the higher education institutions and the people that filled the questionnaire; the second section consists of the questions that determine the factors related with the misuse of information and communication technologies in the universities and the third section is composed by the questions that complements the study. Additionally, the survey was administered to 50 students before the data collection in order to verify if the questions were related with the study objectives.

## 2.2. Factors

Table 1 presents the technological factors considered for the prediction process of university student desertion. Factors such as age, gender, marital status along with technological aspects served as variables of the prediction models.

**Table 1**  
Factors that influence the student university desertion

Factor	Description
Gender	male, female
Age	17 - 40 years
Marital Status	Single, married, divorced, widow
Addiction to Internet	The student that uses internet without academic

	purposes
Addiction to social media	The student that uses social media without academic purposes
Addiction to technology	Technology usage (computer, cellular phones, Tablets) without academic purposes
Des	Intention of the student to drop out college

### 3. Results

The data collected includes 7 factors that relate the inadequate usage or abuse of information and communication technologies with the university desertion and can constitute variability sources in the dropout rates. The factor desertion is considered as the explanatory variable while factors such as age, gender, marital status, addiction to internet and addiction to technologies are considered explanatory variables.

In order to construct the models, data mining algorithms were used via two techniques: decision trees and logistic regression. Logistic regression was applied because it is an analysis method for dichotomic variables and presents higher estimations compared with other methods (Hosmer Jr, Lemeshow, & Sturdivant, 2013). Additionally, decision tree classifier was used because it is a transformation technique for the conversion of explanatory variables. It is easy to interpret the results and provide better precision rates (Márquez-Vera, Morales, & Soto, 2013). Moreover, decision tree classifier uses a methodology that allows to identify the most significant variables in the model and the established values in the tree structure (Thammasiri, Delen, Meesad, & Kasap, 2014).

The experimental process starts with the data parameter adjustment in order to be used in the supervised learning. The preprocessing stage comprises of three activities, integration, cleaning and transformation of the information. This stage was used for anomalies management, atypical value correction and lost values. These activities were performed to improve the variables' characteristics and optimization of the searching process of data mining algorithms.

The technique relief feature selection was applied in the stage of relevant attribute selection. This technique evaluates value of an attribute by repeatedly showing one instance considering the given value of the attribute for the closest instance of the same class and different. The results of the factor selection process are displayed in Table 2.

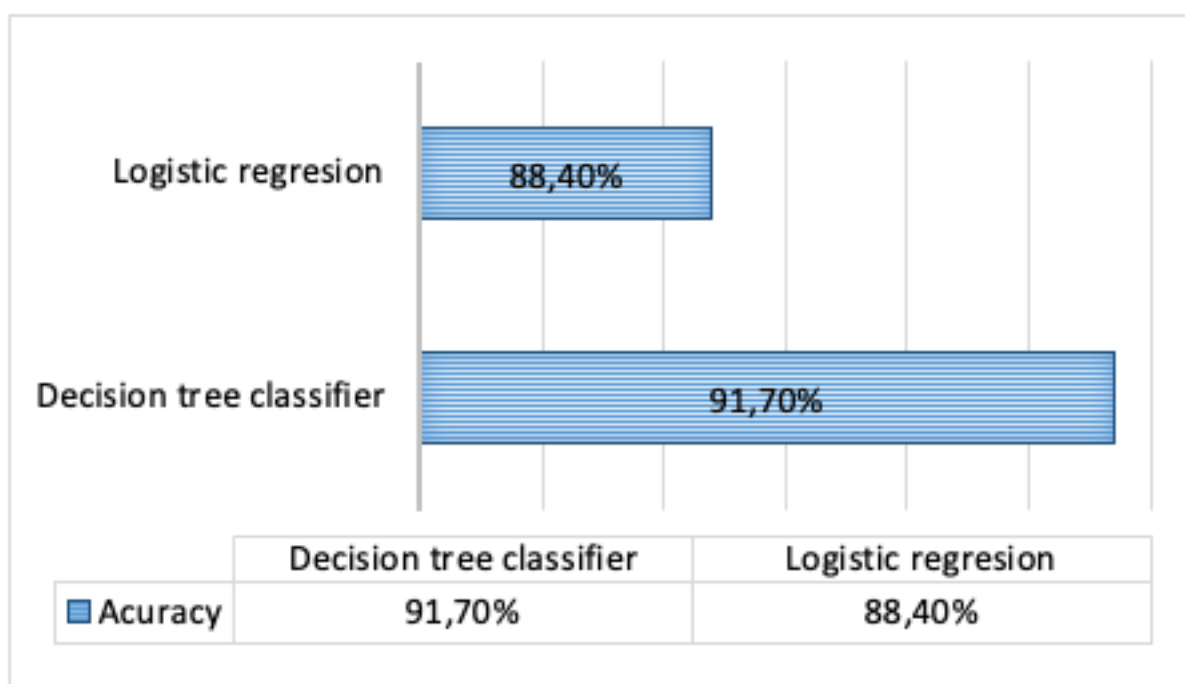
**Table 2**  
Selection of the factors the student university desertion

Factor Attribute Ranking	
<b>Attribute Evaluator (supervised, Class (numeric): 9 DES):</b>	
	Relief Ranking Filter
	Instances sampled: all
Number of nearest neighbors (k): 10	
Equal influence nearest neighbors	
<b>Ranked attributes:</b>	

0.01017	6 Internet
0.00977	1 Gender
0.00844	7 Adic_Tec
0.00752	6 Red_S
0.00625	2 Age
-0.00288	3 Marital Status

Finally, the prediction process of university desertion is established by the application of the factors mentioned above. This process is utilized for inferring objective attributes or data aspects and in the education case is used for detecting the students' behavior (Romero & Ventura, 2013). Furthermore, 10 tests for each logistic regression algorithm and decision tree classifier were performed in order to understand the importance of the factors' characteristics utilized as entry variables for the models. Metric accuracy was employed for determining the precision rate of the desertion prediction. The results show that the method that presents the higher precision is the decision tree classifier with 91.7%, which is presented in Figure 2.

**Figure 2**  
Accuracy of the prediction models



## 4. Conclusions

The article focuses on automatic learning through the use of technological factors that influence in the students' decision to drop out from the university. In this study two techniques were compared: tree decision prediction and logistic regression along with seven factors that served as entry variables to the prediction models. For the university student desertion prediction process, 4 data mining stages were established starting with the preprocessing until the prediction of desertion phase.

The study results show that the tree decision method allows predicting the desertion with a higher level of precision of 91.7%. The establishment of technological factors to predict the university desertion highlights the importance of the study since it was possible to determine the causality relationship of these factors with college drop out.

According to the results presented in Table 2, the factor that has the major influence according to the attribute ranking of factors is the addiction to internet that corresponds to the usage of internet without academic purposes with a value of 0,1017. This result can be

explained to the fact that 43.07% of the surveyed students affirmed using internet without academic purposes for more than 15 hours per week. Therefore, this time spend surfing on the internet can be utilized for academic purposes.

Even though the data collected in this study was large, increasing the sample size and factors could improve the prediction results. Variables such as the number of times students enter in different internet activities such as facebook, instagram and whatsapp can be considered for future research.

---

## **Bibliographic references**

- Argote, I., & Jimenez, R. (2016). Detección De Patrones De Deserción En Los Programas De Pregrado De La Universidad Mariana De San Juan De Pasto, Aplicando El Proceso De Descubrimiento De Conocimiento Sobre Base De Datos (Kdd) Y Su Implementación En Modelos Matemáticos De Predicción. Paper presented at the Congresos CLABES.
- Blau, I. (2011). Application use, online relationship types, self-disclosure, and Internet abuse among children and youth: Implications for education and Internet safety programs. *Journal of Educational Computing Research*, 45(1), 95-116.
- Bustos Andreu, H., & Nussbaum, M. (2009). An experimental study of the inclusion of technology in higher education. *Computer Applications in Engineering Education*, 17(1), 100-107.
- Carbonell, X., Chamarro, A., Griffiths, M., Oberst, U., Cladellas, R., & Talarn, A. (2012). Problematic Internet and cell phone use in Spanish teenagers and young students. *Anales de Psicología/Annals of Psychology*, 28(3), 789-796.
- Díaz Peralta, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. *Estudios pedagógicos (Valdivia)*, 34(2), 65-86.
- Djulovic, A., & Li, D. (2013). Towards freshman retention prediction: a comparative study. *International Journal of Information and Education Technology*, 3(5), 494.
- Gencer, S. L., & Koc, M. (2012). Internet abuse among teenagers and its relations to internet usage patterns and demographics. *Journal of Educational Technology & Society*, 15(2), 25.
- Ghamari, F., Mohammadbeigi, A., Mohammadsalehi, N., & Hashiani, A. A. (2011). Internet addiction and modeling its risk factors in medical students, Iran. *Indian journal of psychological medicine*, 33(2), 158.
- Henríquez, C., & Escobar, R. (2016). Construcción de un modelo de alerta temprana para la detección de estudiantes en riesgo de deserción de la Universidad Metropolitana de Ciencias de la Educación. *Revista mexicana de investigación educativa*, 21(71), 1221-1248.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398): John Wiley & Sons.
- Kittinger, R., Correia, C. J., & Irons, J. G. (2012). Relationship between Facebook use and problematic Internet use among college students. *Cyberpsychology, Behavior, and Social Networking*, 15(6), 324-327.
- Márquez-Vera, C., Morales, C. R., & Soto, S. V. (2013). Predicting school failure and dropout by using data mining techniques. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 8(1), 7-14.
- Méndez-Estrada, V. H., & Barrientos Llosa, Z. (2013). Uso de tecnologías de la informática y comunicación (TIC) para disminuir la deserción de egresados en posgrados semipresenciales. *UNED Research Journal/Cuadernos de Investigación UNED*, 4(2).
- MORENO, C., & MOLINA, Y. (2014). *Evaluación del proceso de retención: desde los que enseñan y aprenden en una educación mediada por las TIC*. Paper presented at the Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación, Buenos Aires, Argentina.
- Murillo, J. (2006). Cuestionarios y escalas de actitudes. *Facultad de Formación de profesorado y Educación. Universidad Autónoma de Madrid*.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.

Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, 41(2), 321-330.

---

1. Professor at the Technical University of Cotopaxi, in the Faculty of Computer Science and Computer Systems, Latacunga, Ecuador. E-mail: [mayra.alban@utc.edu.ec](mailto:mayra.alban@utc.edu.ec)

2. Professor at the National University San Marcos, IA Gropup, Faculty of Postgraduate of Systems and Informatics, Lima, Perú. E-mail: [dmauricios@unmsm.edu.pe](mailto:dmauricios@unmsm.edu.pe)

---

Revista ESPACIOS. ISSN 0798 1015  
Vol. 39 (Nº 52) Year 2018

[Index]

[In case you find any errors on this site, please send e-mail to [webmaster](#)]