

## Analysis of Student Desertion in a Systems and Computing Engineering Undergraduate Program

## Análisis de deserción estudiantil en un programa de pregrado en Ingeniería de Sistemas y Computación

Luis Fernando Castro R.<sup>1</sup> , Esperanza Espitia P.<sup>1</sup> and Sergio Augusto Cardona<sup>1</sup> 

<sup>1</sup>Universidad del Quindío, Armenia, Colombia

lufer@uniquindio.edu.co, eespitia@uniquindio.edu.co, sergio\_cardona@uniquindio.edu.co

(Received: 15 February 2019; accepted: 22 February 2019)

### Abstract

Data mining techniques are mainly focused on supporting the decision makers in a specific organization. Student attrition is a common phenomenon that worries public and private universities, which are affected financially and socially. Several studies have addressed this issue. However, they have mainly focused on academic, social, demographic, and economic aspects. In this paper, we propose a method for analyzing academic desertion in the context of a Systems and Computing Engineering undergraduate program by providing a view of this issue from a KDD (knowledge discovery in databases) perspective and using techniques for identifying students' behavioral patterns. Unlike other proposals, we also consider variables provided by the BADyG test. This proposal is important because it will support higher education institutions in decision-making and creating action plans to reduce the high rate of student attrition.

**Keywords:** Data Mining, Student attrition, KDD, Patterns, CRISP-DM, Analysis.

### Resumen

Las técnicas de minería de datos se enfocan principalmente en apoyar el proceso de toma de decisiones dentro de una organización. La deserción estudiantil es un fenómeno común que agobia a las universidades tanto públicas como privadas, las cuales se afectan de manera social y económica. Diversos estudios se llevaron a cabo en esta área; sin embargo, por lo general se enfocan solo en los aspectos académicos, sociales, demográficos y económicos. Este artículo propone un método para analizar la deserción académica en el contexto de un programa de pregrado en Ingeniería de Sistemas y Computación. Proporciona una vista de esta problemática desde la perspectiva ofrecida por KDD (descubrimiento de conocimiento en bases de datos) y usa técnicas para descubrir patrones de comportamiento asociados con dicha problemática. A diferencia de otros trabajos similares, esta propuesta considera variables planteadas por las pruebas BADyG. Este trabajo proporcionará apoyo al proceso de toma de decisiones y fomentará la creación de planes de acción por parte de las instituciones de educación superior con el propósito de reducir la preocupante tasa de deserción estudiantil.

**Palabras Clave:** Minería de datos, deserción estudiantil, patrones, CRISP-DM, análisis.

## 1. Introduction

Nowadays, student attrition is an issue that worries public and private universities. This problem's causes have not been accurately identified. Besides, the way of using data associated with students in order to generate useful information that can help face this problem is a challenge. According to (Hernán Cáceres & González Cardona, 2011), one of the main difficulties the current educational system faces is desertion. Its accumulated value reaches levels of 45% on a national level; one of its main causes is academic desertion. According to Timarán and Jiménez (2015), the main problem the Colombian higher education system is



Cite this work as Castro L., Espitia E. and Cardona S. (2019). Analysis of Student Attrition in a Systems and Computing Engineering Undergraduate Program Revista Colombiana de Computación, 20(1), 72-82. <http://dx.doi.org/10.29375/25392115.3608>

facing are the high levels of student attrition (Timarán Pereira & Jiménez Toledo, 2015). They argue that the number of students that have completed their higher education is very low, perceiving that most of them abandoned their degrees during the first semester. In addition, half the number of students enrolled in a higher education institution cannot complete their academic cycle. Finally, they state that student attrition was estimated to be 49% in 2004, presenting the following causes: economic and financial constraints, low academic performance, vocational and professional disorientation and difficulties in adapting to a university environment. According to Argote and Jiménez (2016), only one of every two students who enroll in an undergraduate program complete their career (Argote & Jimenez, 2016). The concern is greater if you consider that 39.52% of those who drop out of their schooling argue financial reasons.

The increase in attrition rates has become a problem of interest to higher education institutions and educational authorities. This issue has significant socio-economic consequences. The loss of students causes serious problems for universities, since it makes their sources of income unstable. In addition, student desertion can compromise the future of a country in the medium and long-term, since the accumulation of scientific and technological knowledge is one of the factors that determine the socio-economic development of a nation (Castaño, Gallón, Gómez, & Vásquez, 2008). According to Paramo and Correa (2012), school failure, in any case, is a catastrophe, absolutely devastating on moral, human and social levels, which very often generates exclusions that will mark young people during their adult lives (Paramo & Correa Maya, 1999). On top of that, desertion and abandonment produce uprooting, loneliness, an absence of rites, lack of routines and loss of negotiating capacities with others, as well as social loneliness. Then, the authors explain that desertion is a problem of the educational system that is related to its environments, such as educational environments, family situations, and environmental and cultural requirements that directly affect the deserter.

Several studies based on KDD have been developed (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). González (2011) uses KDD in knowledge extraction from the academic information system's (SIA) database of the University of Caldas in order to calculate academic performance indicators (Hernán Cáceres & González Cardona, 2011). Hernández Cáceres & Gutiérrez (2012), presented that useful knowledge is generated from the large amounts of academic information in the academic registry office in order to find possible causes of the student attrition problem of the autonomous University of Manizales (Hernández Cáceres & Gutiérrez, 2012). The work presented by Argote and Jiménez (2016) contributes to decision-making for reducing student attrition levels in Mariana University's undergraduate programs by applying the KDD process and based on a unified data repository with students' socio-economic, personal, academic and institutional information (Argote & Jimenez, 2016). Azoumana (2013) performs an analysis of student attrition in Simon Bolivar University's systems engineering program. There, the causes of desertion are grouped into variables based on information from the academic registry office in order to establish patterns and support decision making using data mining techniques and the tool for automatic learning and data mining called WEKA (Azoumana, 2013). In this paper, we propose a method for analyzing academic desertion by using techniques, tools and methodologies based on KDD. Unlike previous projects, we are focused on the systems and computer engineering program's students from the University of Quindío. Besides, we include a set of variables provided by the department of statistics, which are the result of a test that assesses different cognitive aspects related to students. These variables are added to the information provided by other information systems, such as the registrar's office and planning department. The data was provided by three sources: information concerning the SPADIES taken from the planning and development office, a BADyG test provided by the department of statistics and the students' personal and academic information, taken from the registrar's and control office of the University of Quindío.

According to the issue of student attrition, some authors say that it is the educational institutions' obligation, especially universities, to establish academic, administrative and adjustment mechanisms to their students' university life for them to overcome the difficulties of the academic programs and successfully culminate their careers (Paramo & Correa Maya, 1999). In addition to the large volume of data available in higher education institutions related to their students, Fayyad et al. (1996) propose that a new generation of computational techniques and tools is required to support extracting useful knowledge from the rapidly growing volumes of data (Fayyad et al., 1996). These techniques and tools are related to an emerging field of knowledge discovery in databases (KDD) and data mining. The information that can be obtained from academic databases will help answer questions, such as: what are the causes of student retention in the university? Why do students drop out? According to Salazar et al. (2004), automatic data mining techniques can be applied to answer these questions and facilitate developing strategies for improving academic processes and educational programs (Salazar, Gosalbez, Bosch, Miralles, & Vergara, 2004). Data mining can offer a great variety of statistical and computational methods to investigate the existence of students' relationships and behavioral

patterns during their first year of university for the issue of student attrition (Hernández Cáceres, 2011).

The paper is organized as follows: Section 2 describes the main concepts for understanding the proposal. In Section 3, we discuss some related work. In Section 4, we describe the proposal together with the methodologies and tools that were used. And finally, some conclusions are presented in Section 5.

## 2. Theoretical Framework

### 2.1 Student Attrition

According to Paramo and Correa (2012), student attrition must be understood as the definitive abandonment of schooling for various reasons and non-continuity of the academic formation of each person who begins their studies, hoping to happily finish their university studies (Paramo & Correa Maya, 1999).

### 2.2 KDD

The term KDD refers to the overall process of discovering useful knowledge from data. In addition, this process focuses on searching for data patterns that are valid, innovative, potentially useful and understandable (Fayyad et al., 1996).

### 2.3 BADyG

The term BADyG refers to a test for assessing different cognitive functions related to the subjects. The theoretical foundation of this test is that intelligence is composed of a set of differentiated capacities instead of a single capacity (Galvis, 2007; Yuste Hernanz & Martínez Arias, 2005). The variables provided by the BADyG test are shown in Table 1.

**Table 1.** Evaluated variables in the BADyG test. Source: (Galvis, 2007; Yuste Hernanz & Martínez Arias, 2005)

Variable	Mean
CI	Coefficient of intelligence
IG	General intelligence
RL	Logic reasoning
Rv	Analog relations
Rn	Numerical series
Re	Logical matrices
Sv	Complete sentences
Sn	Numerical problems
Se	Fit figures
Ma	Auditory memory
Mv	Visual memory
De	Discrimination of differences
RA	Speed
EF	Effectiveness

### 2.4 Data mining

Data mining is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules, allowing, for example, a corporation to improve its marketing, sales, and customer support operations through a better understanding of its customers (Linoff & Berry, 2011).

### 2.5 CRISP-DM

CRISP-DM is a methodology widely used for analyzing large volumes of data and discovering valuable information. According to Chapman et al. (2000), the CRISP-DM methodology consists of six phases: understanding the business, understanding the data, preparing the data, modeling, evaluation and implementation (Chapman et al., 2000).

## **2.6 Rapid Miner**

Rapid Miner is an open-source software (OSS). Rapid Miner provides functionality data preprocessing and visualization, predictive analytics and statistical modeling, in addition to data mining. There are a lot of OSSs in the market but Rapid Miner's user interface and workflow makes it different from other tools (Santhanakumar & Christopher Columbus, 2015).

## **3. Related Work**

Several projects use data mining techniques to identify behavioral patterns that are useful for a particular context. Cruz O. & Ortega C (2008); Eckert & Suénaga (2015) propose collecting statistical data in order to identify behavioral patterns about travelers visiting the city of Pereira, describe their habits and classify those habits with respect to trends. This project was carried out by using the CRISP-DM methodology, the KDD process, and the tool called RapidMiner. The project presented by Hernández Cáceres & Gallego Gallego (2014) generates useful knowledge from historical data analyses of successful and failed projects on an IT outsourcing company by applying data mining techniques, in order to obtain patterns for allowing the organization to make decisions and guide software projects towards success (Hernández Cáceres & Gallego Gallego, 2014). This work was done by using the CRISP-DM methodology, the KDD process and the SPSS (statistical product and service solutions) tool. Sotomonte et al. (2016) use data mining techniques to address the problem of student attrition at the Universidad Distrital – Francisco José de Caldas – in order to determine the causes that lead to student desertion at the university (Sotomonte-Castro, Rodríguez-Rodríguez, Montenegro-Marin, Gaona-García, & Castellanos, 2016). This project is intended to generate a decision tree model by implementing the J48 algorithm, using the WEKA tool to identify such causes. Also, they used the CRISP-DM methodology to develop the project. Finally, the work presented by Vélez Bedoya & Salcedo Toro (2015) analyzes the academic information related to students' academic results and their interactions with the university (Vélez Bedoya & Salcedo Toro, 2015). The authors identify factors that influence student desertion from the computer science degree at Gastón Dachary University in Argentina. This project applies data mining techniques and uses classification algorithms such as decision trees, Bayesian networks and rules. The project is developed under the KDD process and the WEKA tool. These studies use the same technology in the same context as our proposal. However, they have mainly focused on academic, social, demographic, and economic aspects. Unlike such proposals, we also consider variables provided by the BADyG test.

## **4. Our Proposal**

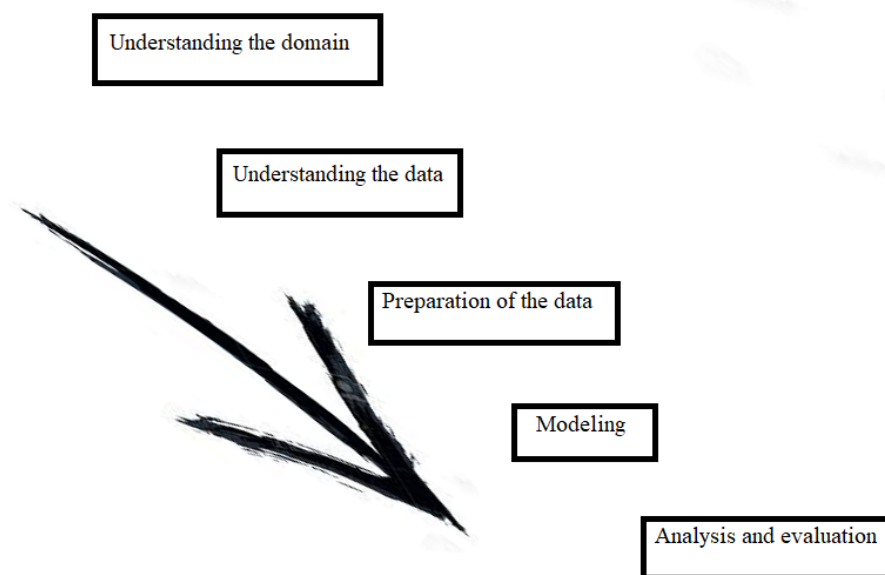
This proposal consists of a method for analyzing academic desertion, starting from the data provided by the University of Quindío related to the students and focusing particularly on the undergraduate program named Systems and Computing Engineering. For this proposal, we intend to work with the CRISP-DM methodology, the KDD process, especially in the data mining stage, and the tool called RapidMiner. The goal of this proposal is to identify some behavioral patterns and relationships between a large number of variables of the students of the University of Quindío. Thus, this work can support decision-making and creating plans focused on addressing problems related to desertion in the University of Quindío's Systems and Computing Engineering program. According to the CRISP-DM methodology, the proposal will include five phases: understanding the domain, understanding the data, preparation of data, modeling, and analyzing and evaluating. Figure 1 illustrates this proposal.

### **4.1 Understanding the Domain**

Developing this phase known as "understanding the domain" in CRISP-DM allows us to obtain a higher level of understanding of the problem related to the case study. In this case, the task consists of consulting research and previous work related to the problem of desertion at the University of Quindío, as well as on a national and international level. Documentation related to knowledge discovery techniques and their application to desertion topics is also reviewed.

### **4.2 Understanding the Data**

In this phase of CRISP-DM, named "understanding the data," the methodology proposes criteria for selecting



**Figure 1.** Phases of the proposal.

the data. This data is obtained and explored in order to identify the elements that allow determining its quality. In this case, the following criteria were defined: student’s data provided by the various departments at the University of Quindío responsible for collecting and storing students’ information. These departments were the registrar and control office, the planning department, and the department of statistics. The source data was provided by three sources: information concerning to the SPADIES taken from the planning and development office, BADyG test provided by the department of statistics and students’ personal and academic information, taken from the registrar and control office of the University of Quindío. The department of statistics at the University of Quindío performed a BADyG test (Galvis, 2007). A sample of the obtained results can be consulted in Figure 2, Figure 3, and Figure 4.

	1998-1	1998-2	1999-1	1999-2	2000-1	2000-2	2001-1	2001-2	2002-1	2002-2	2003-1	2003-2	2004-1	2004-2	2005-1	2005-2	Classification
Example 1	Registered	Registered	Registered				Registered	Registered	Registered	Registered							deserter
Example 2											Registered	Registered	Registered	Registered			deserter
Example 3	Registered	Registered															deserter
Example 4								Registered	Registered								deserter
Example 5											Registered	Registered	Registered	Registered			deserter
Example 6	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered			deserter
Example 7	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered			deserter
Example 8	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered			graduate
Example 9			Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered			graduate
Example 10													Registered	Registered			graduate
Example 11	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered	Registered			graduate
Example 12													Registered	Registered			graduate
Example 13													Registered	Registered	Registered	Registered	active
Example 14													Registered	Registered	Registered	Registered	active
Example 15													Registered	Registered	Registered	Registered	active
Example 16																Registered	active
Example 17																Registered	active
Example 18																Registered	active
Example 19	Registered	Registered	Registered	Registered	Registered	Registered	retired										forced retirement
Example 20	Registered								retired								forced retirement
		Registered															Registered for the period

**Figure 2.** Structure of the information provided by the SPADIES system. Source: (Castro, Espitia & Montaña, 2018).

ESTP_ID	IG	RL	Rv	Rn	Re	Sv	Sn	Se	Ma	Mv	De	RA	EF
412345	97	47	16	15	16	15	12	23	21	18	15	143	68
412340	104	54	20	13	21	18	9	23	20	9	23	164	63
413207	59	32	17	8	7	12	0	15	20	13	11	136	43
415589	67	37	12	14	11	8	7	15	8	3	21	100	67
412350	46	24	12	4	8	11	4	7	10	9	17	120	38
412338	95	48	15	17	16	11	12	24	17	10	26	118	81
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
412360	68	43	11	15	17	5	1	19	24	5	21	111	61

Figure 3. Structure of the information provided by the statistical department of the University of Quindío related to BADyG test.

ESTP_ID	TIPO_DOC	PROG_NOMBRE	CIGE_NOMBRE	GENERO	ASPL_FECHANAMIENTO	VIGENCIA				ESTADO_CIVIL
						2015-2	2016-1	2016-2	2017-1	
408121	CC	INGENIERIA DE SISTEMAS Y COMPUTACION	BOGOTÁ D.C.	M	35090	1				SOLTERO(A)
412333	TI	INGENIERIA DE SISTEMAS Y COMPUTACION	ARMENIA	M	35755	1	1	1	1	SOLTERO(A)
412336	TI	INGENIERIA DE SISTEMAS Y COMPUTACION	ARMENIA	M	35852	1	1	1	1	SOLTERO(A)
412337	CC	INGENIERIA DE SISTEMAS Y COMPUTACION	ARMENIA	M	35462	1	1	1	1	SOLTERO(A)
412338	CC	INGENIERIA DE SISTEMAS Y COMPUTACION	MANIZALES	M	35329	1	1	1	1	SOLTERO(A)
412339	TI	INGENIERIA DE SISTEMAS Y COMPUTACION	ARMENIA	F	35761	1				SOLTERO(A)
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
412342	TI	INGENIERIA DE SISTEMAS Y COMPUTACION	ARMENIA	M	35612	1	1	1	1	SOLTERO(A)

Figure 4. Structure of the information provided by the planning and development department of the University of Quindío related to the personal and academic information of the students.

### 4.3 Preparation of Data

This phase named “preparation of data” in CRISP-DM consists of organizing and debugging the obtained data. This phase is composed of four general tasks and four specific tasks. Task 1 allows us to determine the data that will be included in the study and to exclude low quality data. In this case, the data related to students’ name and address were excluded. Such data was classified as sensitive information. Task 2 performs the data’s cleaning process, starting from the previously defined selection criteria. Task 3 consists of structuring the data. In this step the tables for storing de data were structured. Finally, Task 4 allows us to integrate the data.

In previous phases, we can see several sets of data that were provided by diverse and heterogeneous sources. So, the information presents great difficulties when it comes to adequately integrating it. For example, Table 2 shows the structure and possible values related to the data provided by the planning and development department.

Table 2. Data structure provided by the planning and development department

Attribute	Value
CIVIL STATUS	Single
ARMED_CONFLICT_VICTIM	Not
ES_DISPLACED	Not
ES_DISABLED	Not
STRATUM	4
SECURITY_REGIME	Contributory (E.P.S)
SISBEN_CATEGORY	Don’t apply
AMOUNT_FAMILY_GROUP	4
FAMILY_NUCLEUS	two parents and brothers
CONTRIBUTE_TO_THE_FAMILY	Not
INCOME_QUANTITY	Between 0-1 minimum salary
FINANCING_SOURCE	Income of your parents and / or families, except husband or partner
WORK	Not
MOTIVATION	Because I’m interested in the comprehensive education offered by universities



As we can see, this data, as well as the data provided by the others dependencies, present several inconsistencies related to heterogeneous structure, missing figures, duplicated records, and redundant information, among others.

In general, all data provided by the different sources should be cleaned and integrated. All this process was carried out using a set of macros in Microsoft Excel®. These macros were developed by the authors. Particularly, the information provided by these heterogeneous sources was imported in several excel tables. Then, we used several intermediate dynamic tables to record the partial results. So, each one of these tables was generated with Microsoft Excel® using macros, formulas and SQL sentences. Finally, the data resulting from the previous tasks were combined and integrated.

#### **4.4 Modeling**

In this phase, modeling techniques are selected and applied, and their parameters are calibrated to optimal values. In this case, we use the classification tree for relating the previously selected attributes. These attributes were selected considering the level of incidence on the decision to desert or not to desert: BADyG test parameters, genre, age, marital status, victim of conflict, displaced, disabled, and stratum, etc. These parameters can be consulted in Figure 5a, 5b.

#### **4.5 Analysis and Evaluation**

In this phase the obtained model (or models) are more thoroughly evaluated and the steps executed to construct the model are reviewed. Consequently, we presented the results once the information was modeled, analyzed and evaluated. These results are illustrated in Figures 6 and 7.

The decision tree in Figure 6 shows that attributes and relationships that encourage a student to desert the systems and computing engineering program are the following:

The number of family members, age, regime, whether they contribute or not, and stratum. According to Figure 6, students who desert for financial reasons have between the age of 21 and 23 and have under 5 family members. Another interesting situation is related to the students who have between 1 and 3 family members, stratum under 5, whose regime is contributory (EPS, Empresa Prestadora de Salud) and have between the age of 19 and 20. The same occurs with previous students whose regime is subsidized and have 4 family members.

Regarding the BADyG test, the most significant attributes that encourage students to desert this academic program are the following:

Students who have a score of less than or equal to 57.5 in IG (General intelligence) have a high probability of deserting. Moreover, students with high probability for deserting have a score of greater than 57.5 in IG (General intelligence), a score of less than or equal to 18.5 in Mv (Visual memory), a score of under 82.5 in EF (Effectiveness), a score of greater than 58.5 in IG (General intelligence) and 3 family members.

### **5. Conclusions**

This paper proposed a method for analyzing student attrition in the Systems and Computing Engineering program at the University of Quindío. Therefore, an explanation of the process was described using the phases provided by the CRISP-DM methodology. This methodology favors obtaining results because it allows us to completely understand the problem in terms of business and its meaning in terms of data mining.

The advantage of this project is that it considers a group of aspects related to intelligence tests. This was achieved by analyzing the data obtained from the BADyG (Battery of General and Differential Aptitudes) tests. This project's results reflect the university's need to establish action plans that improve students' performance, such as extracurricular advice and accompaniment to strengthen the flaws that were found. This strategy can even be applied in other universities.

Some inconveniences were found when applying university policies to determine the status of students who have deserted. They were not very clear or in accordance with reality. Consequently, it is necessary to develop additional studies to improve the classification of students who have deserted in future projects.

#### **Declaration of conflicting interests**

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

ESTP_ID	IG	RL	Rv	Rn	Re	Sv	Sn	Se	Ma	Mv	De	RA	EF
412345	97	47	16	15	16	15	12	23	21	18	15	143	68
412340	104	54	20	13	21	18	9	23	20	9	23	164	63
413207	59	32	17	8	7	12	0	15	20	13	11	136	43
415589	67	37	12	14	11	8	7	15	8	3	21	100	67
412350	46	24	12	4	8	11	4	7	10	9	17	120	38
412338	95	48	15	17	16	11	12	24	17	10	26	118	81
412334	73	44	14	16	14	12	2	15	23	12	25	104	70
412379	114	66	18	23	25	14	11	23	24	18	24	135	84
412352	57	26	12	8	6	12	7	12	12	11	22	127	45
412335	74	40	18	10	12	16	7	11	9	12	20	166	45
413208	85	50	23	15	12	18	7	10	21	10	17	143	59
412384	82	43	15	17	11	17	10	12	23	24	21	170	48
412456	101	66	21	22	23	13	15	7	18	19	15	135	75
412360	68	43	11	15	17	5	1	19	24	5	21	111	61
412336	71	36	13	13	10	10	7	18	26	17	18	114	62
412455	81	43	12	16	15	11	10	17	17	15	22	99	82
412344	26	8	3	1	4	8	2	8	8	12	14	92	28
412337	90	51	15	17	19	11	10	18	22	18	18	98	92
412347	77	39	18	8	13	17	2	19	17	11	28	159	48
415233	97	54	16	23	15	13	16	14	15	22	16	150	65
415232	65	33	8	16	9	12	8	12	12	17	24	127	51
412339	77	47	15	12	20	9	10	11	13	16	22	115	67
412351	71	38	17	9	12	14	4	15	14	16	20	129	55
412342	83	42	12	14	16	11	9	21	12	8	23	100	83
412454	60	34	12	11	11	10	7	9	6	13	16	112	54
412333	96	48	12	17	19	17	6	25	27	27	22	124	77
415308	58	23	11	0	12	15	6	14	26	14	17	93	62
412373	65	33	8	11	14	15	4	13	21	15	19	115	57
412459	85	53	18	14	21	12	5	15	20	15	22	140	61
412458	50	30	10	11	9	8	5	7	13	12	14	109	46

Figure 5a. Consolidated final data.

ESTP_ID	ESTADO_CIV	VICTIMA	COMFLI	ES_DESPLAZA	ES_DISCAPACIT	ESTRATO
412345	SOLTERO(A)	NO	NO	NO	NO	2
412340	SOLTERO(A)	NO	NO	NO	NO	3
413207	SOLTERO(A)	NO	NO	NO	NO	2
415589	SOLTERO(A)	NO	NO	NO	NO	1
412350	SOLTERO(A)	NO	NO	NO	NO	1
412338	SOLTERO(A)	NO	NO	NO	NO	6
412334	SOLTERO(A)	NO	NO	NO	NO	1
412379	SOLTERO(A)	NO	NO	NO	NO	3
412352	SOLTERO(A)	NO	NO	NO	NO	3
412335	SOLTERO(A)	NO	NO	NO	NO	2
413208	SOLTERO(A)	NO	NO	NO	NO	1
412384	SOLTERO(A)	NO	NO	NO	NO	2
412456	SOLTERO(A)	NO	NO	NO	NO	2
412360	SOLTERO(A)	NO	NO	NO	SI	1
412336	SOLTERO(A)	NO	NO	NO	NO	1
412455	SOLTERO(A)	NO	NO	NO	NO	2
412344	SOLTERO(A)	NO	NO	NO	NO	3
412337	SOLTERO(A)	SI	NO	NO	SI	2
412347	SOLTERO(A)	NO	NO	NO	NO	1
415233	SOLTERO(A)	NO	NO	NO	NO	2
415232	SOLTERO(A)	NO	NO	NO	NO	3
412339	SOLTERO(A)	NO	NO	NO	SI	1
412351	CASADO(A)	NO	NO	NO	NO	1
412342	SOLTERO(A)	NO	NO	NO	NO	3
412454	SOLTERO(A)	NO	NO	NO	NO	2
412333	SOLTERO(A)	NO	NO	NO	NO	2
415308	SOLTERO(A)	NO	NO	NO	NO	3
412373	SOLTERO(A)	NO	NO	NO	NO	1
412459	SOLTERO(A)	NO	NO	NO	NO	2
412458	SOLTERO(A)	NO	NO	NO	NO	1

Figure 5b. Consolidated final data.



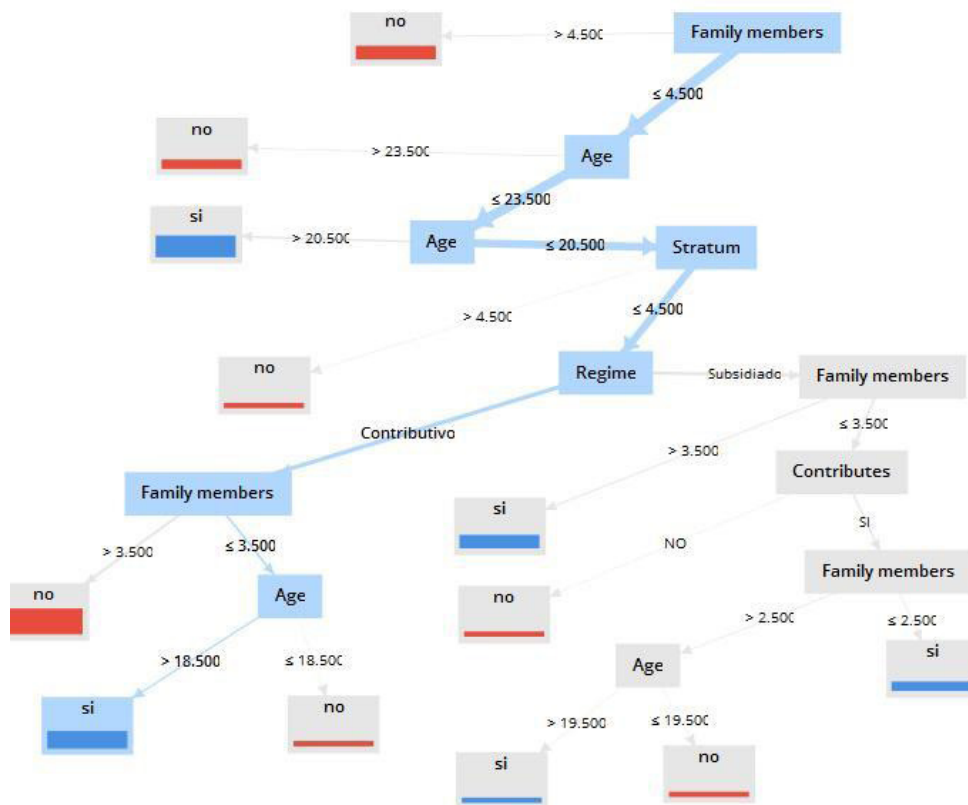


Figure 6. Attrition analysis for the Systems and Computing Engineering program.

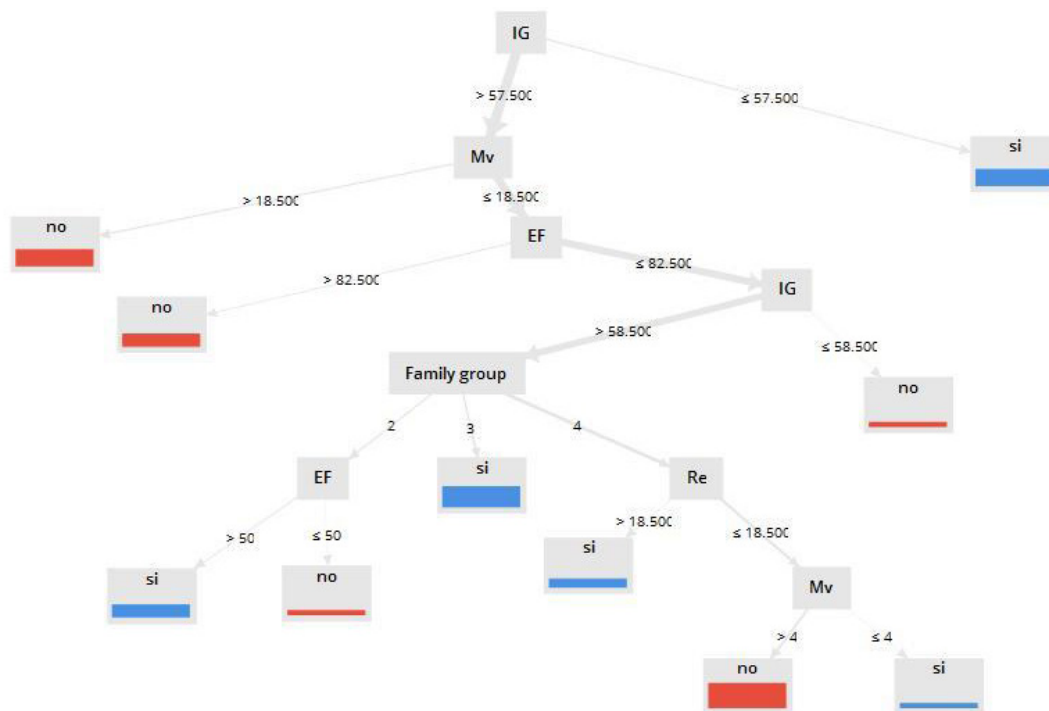


Figure 7. BADyG test for the systems and computing engineering program

## References

- Argote, I., & Jiménez, 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 KDD y su implementación en modelos matemáticos de predicción. In *Conferencia Latinoamericana sobre Abandono en la Educación Superior. Ponencias de Congresos CLABES* (pp. 1–7). Retrieved from <http://revistas.utp.ac.pa/index.php/clabes/article/view/991>
- Azoumana, K. (2013). Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos. *Revista Pensamiento Americano*, 6(10), 41–51.
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, S. (2008). Análisis de los factores asociados a la deserción estudiantil en la Educación Superior: un estudio de caso. *Revista de Educación*, 255–280.
- Castro, L. F., Espitia E., & Montilla A. (2018). Applying CRISP-DM in a KDD process for the analysis of student attrition. *Communications in Computer and Information Science*. Springer, 885, 386-401.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.
- Cruz O., D., & Ortega C., J. (2008). Análisis de la deserción estudiantil en la facultad de Ciencias Exactas y Naturales de la Universidad de Nariño desde la cohorte 2001-2 hasta la cohorte 2006-2 utilizando el sistema SPADIES. Retrieved from <http://sired.udenar.edu.co/214/>
- Eckert, K. B., & Suénaga, R. (2015). Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos. *Formación Universitaria*, 8(5), 03-12. <https://doi.org/10.4067/S0718-50062015000500002>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27–35.
- Galvis, D. (2007). *Estudio sobre la deserción estudiantil en la Universidad del Quindío*. Retrieved from <https://catalogo.uniquindio.edu.co/cgi-bin/koha/opac-detail.pl?biblionumber=33407>
- Hernán Cáceres, J., & González Cardona, J. C. (2011). *Sistema de apoyo para la acreditación de la calidad de programas académicos de la Universidad de Caldas, aplicando técnicas en minería de datos*. Universidad Autónoma de Manizales. Retrieved from <http://repositorio.autonoma.edu.co/jspui/handle/11182/38>
- Hernández Cáceres, J. (2011). Descubrimiento de conocimiento en la base de datos académica de una institución de educación superior usando redes neuronales. *Vector*, 7–19.
- Hernández Cáceres, J., & Gallego Gallego, M. (2014). *Descubrimiento de conocimiento en una empresa de outsourcing de TI de la ciudad de Medellín aplicando técnicas de minería de datos que permita identificar potencialidades en el éxito de los proyectos de desarrollo de software*. Universidad Autónoma de Manizales. Retrieved from <http://repositorio.autonoma.edu.co/jspui/handle/11182/51>
- Hernández Cáceres, J., & Gutiérrez, J. E. (2012). *Descubrimiento de conocimientos en la base de datos académica de la Universidad Autónoma de Manizales aplicando redes neuronales*. Universidad Autónoma de Manizales. Retrieved from <http://repositorio.autonoma.edu.co/jspui/handle/11182/39>
- Linoff, G., & Berry, M. (2011). Why and What is Data Mining? In *Data Mining Techniques*.
- Paramo, G. J., & Correa Maya, C. A. (1999). Deserción estudiantil universitaria. Conceptualización. *Revista Universidad EAFIT*, 35(114), 65–78. Retrieved from <http://publicaciones.eafit.edu.co/index.php/revista-universidad-eafit/article/view/1075>
- Salazar, A., Gosalbez, J., Bosch, I., Miralles, R., & Vergara, L. (2004). A case study of knowledge discovery on academic achievement, student desertion and student retention. In *ITRE 2004. 2nd International Conference Information Technology: Research and Education* (pp. 150–154). IEEE. <https://doi.org/10.1109/ITRE.2004.1393665>
- Santhanakumar, M., & Christopher Columbus, C. (2015). Web Usage Based Analysis of Web Pages Using RapidMiner. *WSEAS Transactions on Computers*, 14, 455–464.
- Sotomonte-Castro, J. E., Rodríguez-Rodríguez, C. C., Montenegro-Marin, C. E., Gaona-García, P. A., & Castellanos, J. G. (2016). Hacia la construcción de un modelo predictivo de deserción académica basado en técnicas de minería de datos - Towards the construction of a predictive model of academic desertion based on data mining techniques. *Revista Científica*, 3(26), 35. <https://doi.org/10.14483/23448350.11089>

Timarán Pereira, S. R., & Jiménez Toledo, J. (2015). Extracción de perfiles de deserción estudiantil en la institución universitaria CESMAG. *InvestigiumIre: Ciencias Sociales y Humanas*, 6(1), 30–44. <https://doi.org/10.15658/CESMAG15.05060103>

Vélez Bedoya, J. I., & Salcedo Toro, D. F. (2015). *Tendencias y características de los viajeros que visitan la ciudad de Pereira por medio de técnicas de minería de datos*. Universidad Autónoma de Manizales. Retrieved from <http://repositorio.autonoma.edu.co/jspui/handle/11182/59>

Yuste Hernanz, C., & Martínez Arias, M. del R. (2005). *BADyG S batería de aptitudes diferenciales y generales*. (CEPE, Ed.).

## **Sobre los autores**

### **Luis Fernando Castro Rojas**

Ingeniero de Sistemas. Magister en Ingeniería de Sistemas, Universidad de los Andes. Doctor en Ingeniería, Universidad Nacional. Profesor Titular programa Ingeniería de Sistemas y Computación, Universidad del Quindío.

### **Esperanza Espitia Peña**

Ingeniera de Sistemas. Magister en Ingeniería de Sistemas, Universidad EAFIT. Profesora Asociada programa Ingeniería de Sistemas y Computación, Universidad del Quindío.

### **Sergio Augusto Cardona Torres**

Ingeniero de Sistemas. Magister en Ingeniería, Universidad EAFIT. Doctor en Ingeniería, UPB. Profesor Asociado Ingeniería de Sistemas y Computación, Universidad del Quindío.