

# Engineering Students' Academic Performance Prediction using ICSES Test Scores and Demographic Data

Sandra Merchán-Rubiano<sup>1</sup>, Adán Beltrán-Gómez<sup>2</sup>, Jorge Duarte-García<sup>3</sup>

<sup>1</sup> M.Sc. *Virtual Learning Environments. Teaching Researcher, OSIRIS Research Group.*  
E-mail: [smerchanr@unbosque.edu.co](mailto:smerchanr@unbosque.edu.co)

<sup>2</sup> M.Sc. *Information and Communication Sciences. Teaching Researcher, GITIS Research Group*

<sup>3</sup> Student, OSIRIS Research Group  
Universidad El Bosque, Bogotá, Colombia

---

Received date: August 16, 2016

Accepted date: December 10, 2017

**How to cite this article:** S. Merchán-Rubiano, A. Beltrán-Gómez, J. Duarte-García, "Engineering Students Academic Performance Prediction using ICSES Test Scores and Demographic Data", *Ingeniería Solidaria*, vol. 13, n.º 21, pp. 53-61, January 2017. doi: <http://dx.doi.org/10.16925/in.v13i21.1729>

---

**Abstract.** *Introduction:* This paper is part of a research project that aims to construct a predictive model for students' academic performance, as result of an iterative process of experimentation and evaluation of the pertinence of some data mining techniques. *Methodology:* This paper was written in 2016 in the Universidad El Bosque, Bogotá, Colombia, and presents a comparative analysis of the performance and relevance of the J48 and Random Forest algorithms, in order to identify the most influential demographic and ICSES score variables, as well as the classification rules, to predict the first year academic performance of the Engineering Faculty students, in Universidad El Bosque, Bogotá, Colombia. *Results:* The analysis process was carried out on 7,644 students' records, and it was developed in two phases. Firstly, the data needed to feed the mining process was extracted and prepared. Secondly, the data mining process itself was implemented through preprocessing data and executing the classification algorithms available in Weka. Some significant variables and rules to predict academic performance are found, according to the studied population characteristics. *Conclusions:* The academic risk seen as the cause of the desertion phenomenon must be studied as a phenomenon itself. Establishing its causes facilitates the creation of preventive strategies for the accompaniment of students through their process, aimed to mitigate the risk of both phenomena.

**Keywords:** academic performance prediction, academic risk prevention, data mining, J48, Random Forest.



## LA PREDICCIÓN DEL DESEMPEÑO ACADÉMICO DE LOS ESTUDIANTES DE INGENIERÍA USANDO LOS RESULTADOS DE LA PRUEBA ICFES Y DATOS DEMOGRÁFICOS

**Resumen.** *Introducción:* este trabajo forma parte de un proyecto de investigación cuyo objetivo es construir un modelo predictivo para el rendimiento académico de los alumnos, como resultado de un proceso iterativo de experimentación y evaluación de la pertinencia de algunas técnicas de minería de datos. *Metodología:* este trabajo fue escrito en el 2016 en la Universidad El Bosque, Bogotá, Colombia y presenta un análisis comparativo del desempeño y relevancia de los algoritmos J48 y Random Forest para identificar las variables más influyentes, demográficas y de resultados del ICFES, así como las reglas de clasificación, para predecir el desempeño académico del primer año de los estudiantes de la Facultad de Ingeniería, en la Universidad El Bosque, Bogotá, Colombia. *Resultados:* el proceso de análisis se realizó con 7644 expedientes de estudiantes, y se desarrolló en dos fases. En primer lugar, se extrajeron y prepararon los datos necesarios para alimentar el proceso de minería de datos. En segundo lugar, se implementó el proceso de minería de datos mediante el pre-procesamiento de datos y la ejecución de los algoritmos de clasificación disponibles en Weka. Se hallaron algunas variables significativas y reglas para predecir el rendimiento académico, según las características de la población estudiada. *Conclusiones:* el riesgo académico, visto como la causa del fenómeno de deserción, debe ser estudiado como un fenómeno en sí mismo. Establecer sus orígenes facilita la creación de estrategias preventivas para el acompañamiento de los estudiantes durante su proceso, que mitiguen el riesgo de ambos fenómenos.

**Palabras clave:** predicción de rendimiento académico, prevención de riesgos académicos, minería de datos, J48, Random Forest.

## A PREDIÇÃO DO DESEMPENHO ACADÊMICO DOS ESTUDANTES DE ENGENHARIA USANDO OS RESULTADOS DA PROVA DO ICFES E DADOS DEMOGRÁFICOS

**Resumo.** *Introdução:* este trabalho faz parte de um projeto de pesquisa cujo objetivo é construir um modelo preditivo para o rendimento acadêmico dos alunos, como resultado de um processo iterativo de experimentação e avaliação da pertinência de algumas técnicas de mineração de dados. *Metodologia:* este trabalho foi escrito em 2016 na Universidad El Bosque, Bogotá, Colômbia, e apresenta uma análise comparativa do desempenho e da relevância dos algoritmos J48 e Random Forest para identificar as variáveis mais influentes, demográficas e de resultados do icfes, bem como as regras de classificação para prever o desempenho acadêmico do primeiro ano dos estudantes da Faculdade de Engenharia, na Universidad El Bosque, Bogotá, Colômbia. *Resultados:* o processo de análise foi realizado com 7644 históricos de estudantes e desenvolveu-se em duas fases. Em primeiro lugar, extraíram-se e prepararam-se os dados necessários para alimentar o processo de mineração de dados. Em segundo lugar, implantou-se o processo de mineração mediante o pré-processamento de dados e a execução dos algoritmos de classificação disponíveis em Weka. Constataram-se algumas variáveis significativas e regras para prever o rendimento acadêmico, segundo as características da população estudada. *Conclusões:* o risco acadêmico, visto como a causa do fenômeno de deserção, deve ser estudado como um fenômeno em si. Estabelecer as origens facilita a criação de estratégias preventivas para o acompanhamento dos estudantes durante seu processo, que diminuam o risco de ambos os fenômenos.

**Palavras-chave:** predição de rendimento acadêmico, prevenção de riscos acadêmicos, mineração de dados, J48, Random Forest.



## 1. Introduction

In the Colombian university context, student desertion, understood as an educative phenomenon, has been studied in order to identify which academic and demographic variables are the most influential. National government studies [1] have found that the most influential demographic and social variables are owning a home, the family's income level, number of siblings, the student's mother education, age, gender and job status. The most influential academic variables are repetition rate (as evidence of low performance), career choice and the ICFES test scores<sup>1</sup> [2]. In those studies and in others conducted by some Colombian universities [3], it has been found that low academic performance in the first year of studies is one of the most important factors that impacts student desertion [4-7]. Nonetheless, academic performance has not been studied as a phenomenon itself. Taking into account other research projects in the international context [8-11], this paper is part of a research project [12-14] where data mining techniques are used in order to predict academic performance, identifying its causes based on the demographic and ICFES test scores information collected during each student's enrollment. This work presents the predictive models for academic performance resulted from J48 and Random Forest classification algorithms, and a comparative analysis of their performance.

## 2. Methodological Development

The process was composed of two typical data mining phases: The first one named "Data extraction and preparation", is composed by the extraction, comprehension and cleaning tasks on the dataset that will feed the model; in addition to modelling the prediction problem. The second phase named "Data mining" consists of the loading, preprocessing and application of the classifier algorithms to the dataset, which gives the results, analysis and interpretation.

<sup>1</sup> The ICFES test is a test given to graduating high school students to measure their performance in different knowledge areas, and is often required by universities in Colombia as part of the admittance process. It is comparable to the SAT test for college admissions in the USA.

## 2.1. Data extraction and preparation

All the available Engineering Programs students' demographic and ICFES scores data were extracted from the University's information systems databases, from 2005 to 2015. A database with 7,644 students' individual records was obtained. Table 1 shows the extracted attributes directly from the database, sorted by category.

**Table 1.** Defined instances and attributes

Students' academic data attributes	Students' demographic data attributes
Average ICFES test score, ICFES subject test scores (biology, math, philosophy, physics, history, chemistry, language, geography and English –Test Type 1–, critical lecture, social and civic competences, natural science, quantitative reasoning –Test Type 2–), students' high school, year and period of enrollment, whether the student has received academic incentives, class schedule type (day or evening classes), first semester grades (6 subjects) and second semester grades (6 subjects)	Date of birth, gender, residence distance to University (calculated from home address), city of origin, social class (ranked 1-6 according to national legislation), marital status, job status, whether the student registers a mother, mother's education, mother's job, mother's city of origin, whether the student registers a father, father's job, father's education, father's city of origin, number of siblings.

Compiled by the authors

Based on a context-sensitive analysis, data was cleaned and prepared through a process that included the following tasks:

- Students' age was calculated from their birthdate.
- Attributes like father and mother occupation were discretized into four values: employed, independent, retired and unemployed. The same task was done with the students and their parents' city of birth (principal, intermediate and small city), and each academic program's courses according to the disciplinary area (basic, professional and complimentary).
- All data having consistency problems were cleaned.
- The enrollment date was classified in two types (1 for the first semester of the year and 2 for the second one).

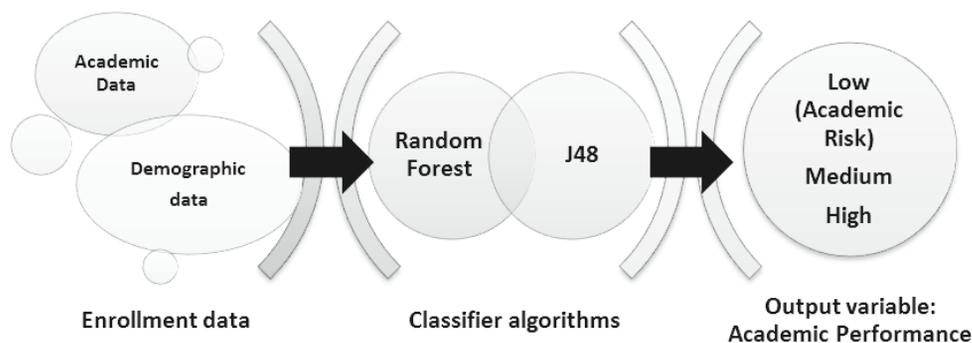
- All attributes with empty or “dirty” records were removed, including whether the student received academic incentives and the name of their high school, among others.
- All records with no ICFES test score data were removed.
- The total dataset was split in two subsets according to the type of ICFES test taken by the students at the end of their basic education (the first was composed by subjects and applied before the second semester of 2014, and the second is composed by competencies and has been applied after the second semester of 2014). Table 2 shows the final data subsets.

**Table 2.** Final data subsets to construct the models

	Type 1 Test	Type 2 Test
Valid instances	4,274	386
Definitive attributes	29	27
Demographic attributes	15	15
Academic attributes	14	12

Source: Compiled by the authors

In the last step of this phase, the prediction problem to be solved was modeled, identifying the input variables, the data mining algorithms and the output variables. Figure 1 shows the prediction problem to resolve:



**Figure 1.** Prediction problem model

Source: Compiled by the authors

To select the data mining algorithms, a preliminary experiment was performed, randomly splitting the total dataset into four subsets and identifying which of the algorithms provided by Weka

[15] could have a better accuracy for the model construction. Table 3 summarizes the results of this experiment.

**Table 3.** Classification confidence percentages of Weka’s classifier algorithms

Algorithm	Experiment 1	Experiment 2	Experiment 3	Experiment 4
<b>Random Forest</b>	<b>68.74</b>	<b>69.64</b>	<b>69.83</b>	<b>69.76</b>
Bayes Net	53.92	63.74	55.43	61.15
Naive Bayes	59.61	59.02	59.13	54.34
<b>J48</b>	<b>68.88</b>	<b>71.23</b>	<b>69.75</b>	<b>69.32</b>
J48 consolidated	51.69	62.31	51.97	41.88

Source: Compiled by the authors

## 2.2. Data Mining

The data was loaded into Weka so it could be mined; this was done by assembling two ARFF files using the text file containing the mining datasets. The dataset was preprocessed according to which

data mining tasks were to be performed. Weka allowed the filtering and transformation of the loaded dataset using RemoveMisClassified and Resampling algorithms; also, attributes from the loaded dataset can be selectively removed.

**Table 4.** Data Preprocessing resulting instances

Test type	Predicted academic performance	Original dataset	RemoveMisclassified	Resampling
Type 1 (before 2014-2)	Low	1,163	402	1,180
	Medium	2,548	2,236	2,546
	High	563	131	548
Type 2 (after 2014-2)	Low	78	32	79
	Medium	232	183	229
	High	76	38	78

Source: Compiled by the authors

The RemoveMisClassified preprocessing filter was applied for removing attributes that don't have relevant influence in the prediction, bearing in mind that the purpose of the study is to identify the most influencing attributes and the relationships between them. Taking into account that the obtained accuracy rates for predicting academic performance vary significantly between output classes, as a result of the imbalanced dataset (most of the instances belonged to the medium academic

performance class), the Re-sampling filter was applied, which oversamples the minority class and undersamples the majority class, obtaining a more balanced distribution of data.

## 3. Results and Analysis

Finally, the algorithms were executed resulting in the classification accuracy rates shown in table 5.

**Table 5.** Algorithms' precision by type of test

	Predicted Academic Performance	Random Forest		J48	
		RemoveMisClassified data	Resampled data	RemoveMisClassified data	Resampled data
ICFES Test	General	85.3	60.6	<b>95.84</b>	68.1
	Low	88.9	70.1	<b>89.3</b>	62
	Medium	85.1	60.3	<b>98</b>	71.2
	High	<b>100</b>	<b>100</b>	80	58.4
Saber 11 Test	General	92	74.9	<b>93.9*</b>	77.4
	Low	94.7	83.9	<b>85*</b>	78.4
	Medium	91.8	73.4	<b>96.6*</b>	77.7
	High	92.3	76.9	<b>75.8*</b>	75.7

\*Binary splits parameter was enabled, and unpruned parameter was disabled

Source: Compiled by the authors

Observing the accuracy rates obtained by both classifier algorithms after preprocessing, it is possible to state that academic risk performance (or low performance) and average performance (medium performance) prediction for those students who took the Type 1 ICFCES test is more accurate using the J48 algorithm and the RemoveMisClassified filter. However, the results obtained by the resampled dataset cannot be completely discarded, because when data is collected (as a student enrolls), the removed attributes remain and the accuracy rates are acceptable. Regarding high academic performance, the algorithm with the best results was Random Forest, which could be a consequence of the level of cleaning achieved in the data subset for this class.

Regarding the academic performance prediction for the group of students who took the ICFCES test after 2014's second semester, although the model was constructed with a relatively small number of instances, high accuracy rates were obtained for the three types of performance with the J48 algorithm. This allows for a high confidence in the rules obtained after preprocessing. Table 6 shows predictions made by both algorithms for the Type 1 ICFCES test dataset, and table 7 shows predictions made for the Type 2 ICFCES test dataset, in the form of confusion matrices.

**Table 6.** Confusion matrices by algorithm for test Type 1 (before 2014-2)

	Random Forest						J48					
	RemoveMisClassified			Resampled data			RemoveMisClassified			Resampled data		
Actual class	L	M	H	L	M	H	L	M	H	L	M	H
Low	<b>136</b>	266	0	<b>68</b>	1,112	0	346	50	6	567	579	4
Medium	17	<b>2,219</b>	0	29	<b>2,517</b>	0	53	2,158	25	300	2,123	123
High	0	124	7	0	542	<b>6</b>	8	26	97	47	281	220

Source: Compiled by the authors

**Table 7.** Confusion matrices by algorithm for test Type 2 (after 2014-2)

	Random Forest						J48					
	RemoveMisClassified			Resampled data			RemoveMisClassified			Resampled data		
Actual class	L	M	H	L	M	H	L	M	H	L	M	H
Low	<b>18</b>	14	0	<b>26</b>	49	4	<b>32</b>	0	0	<b>40</b>	35	4
Medium	1	<b>179</b>	3	5	<b>213</b>	11	1	<b>182</b>	0	10	<b>206</b>	13
High	0	2	<b>36</b>	0	28	<b>50</b>	0	0	<b>38</b>	1	24	<b>53</b>

Source: Compiled by the authors

From the above confusion matrices and accuracy rates, it could be said that reliable predictive models were reached for both data subsets. Although good accuracy rates were obtained applying the resampling filter, the best results were obtained just by removing the misclassified instances.

Table 8 shows the most important rules obtained with the best accuracy rates, corresponding to the J48

algorithm, after removing the misclassified instances, for the Type 1 ICFCES test dataset (first row) and for the Type 2 ICFCES test dataset (second row). The most relevant rules to predict student's academic performance for the Engineering Faculty at Universidad El Bosque were determined, based on demographic and academic enrollment data.

**Table 8.** Rules obtained by J48 algorithm for the Type 1 ICFES test dataset

Predicted academic performance level		
Low (risk)	Medium (average)	High (outstanding)
(ICFES test general score <= 46.81 and gender= female and daytime classes) and (social stratum= 4 and job status= none) or (social stratum= 3 or 2, and enrollment period= 1) (ICFES test general score <= 46.81 and gender= male and residence_city_size = 1) or (hometown size= 4 or hometown size= 3) and history test score <= 51.16) (ICFES test general score <= 46.81 and gender= male and residence_city_size = 2, 3, 4 and job_status= none) (46.81 > ICFES test general score <= 59,48 and chemistry_score<= 44,28 and age= young_adult or (age!= young_adult and physics_score <= 54)) (46.81 > ICFES test general score <= 59,48 and chemistry_score > 44,28 and math_score <= 70 and gender= female or (gender= male and daytime_classes and father_hometown_size= 4) ) (46.81 > ICFES test general score <= 59,48 and chemistry_score > 44,28 and math_score > 70 and gender= female)	(ICFES test general score <= 46.81 and gender= female and (daytime classes) or (evening classes and enrollment period= 1 and social stratum= 3)) (ICFES test general score <= 46.81 and gender= male and residence city size= 1 and father hometown size= 1, 2, 3) (ICFES test general score <= 46.81 and gender= male and residence city size!= 1 or (daytime classes and job_status= none) or (evening classes)) (46.81 > ICFES test general score <= 59,48 and chemistry_score <= 44,28 or (gender= female and daytime classes) or (gender= male and geographics_score > 57.42)) (46.81 > ICFES test general score <= 59,48 and chemistry_score > 44,28 or (daytime classes) or (evening classes or (gender= female and math_score > 41,64) or (gender= male)) (46.81 > ICFES test general score <= 59,48 and chemistry_score > 44,28 or (gender= female and math score <= 70 and daytime classes) or (gender= male and hometown_size = 4 or (job_status= none and philosophy_score > 43,6) or (job_status!= none) or (father hometown size!= 4 and biology_score > 46.64) (46.81 > ICFES test general score <= 59,48 and chemistry_score > 44,28 gender= male and math score > 70) (ICFES test general score > 59,48 and job_status= none and daytime classes and english_score <= 79,5 and mother's_education = Univ.)	(46.81 > ICFES test general score <= 59,48 and chemistry_score > 44,28 and evening classes and gender = female and job_status!= none) (46.81 > ICFES test general score <= 59,48 and chemistry_score > 44,28 and math_score > 70) (ICFES test general score > 59,48 and job_status= none and daytime classes or (english_score <= 79,5 and mother's_education= Univ.) or (english_score > 79,5 and gender= female) (ICFES test general score > 59,48 or (job_status= none and father hometown size=2,3,4) or (job_status!= none and math_score > 60)
Test general score <= 65 and gender= male and test mathematics <= 56	(Test general score <= 65, gender= female) or (test general score <= 65, test general score >= 56, gender= male) or (test general score > 65, test quantitative reasoning <= 62) or (test general score 65, test quantitative reasoning > 62, evening classes)	(Test general score 65, test quantitative reasoning > 62, daytime classes)

Source: Compiled by the authors

In relation with academic risk prediction for the first dataset, this one impacts more on students with a ICFES test general score below 46.8, physics and chemistry scores below 44 and 54 respectively. This impacts more on students with a medium social stratum who are not working at the time of enrollment. When the ICFES test general score rises and the job status is not working, performance is

favored. The rules found to predict academic performance for those students that took the Type 2 test, indicate that the most influential variables are the ICFES test general score, gender, daytime/evening classes and the quantitative reasoning scores obtained on the test. The low performance is mainly influenced in the general test score, male gender, and math score below 56.

## 4. Conclusions and Discussion

Once more it is proven that data mining techniques for predicting modelling are useful in educational contexts. Particularly, at the Engineering Faculty of Universidad El Bosque, relevant and suitable attributes have been found for the studied population. Academic risk as the cause of the desertion phenomenon must be studied as a phenomenon itself. Establishing its causes facilitates the creation of preventive strategies for the accompaniment of students through their process, aimed to mitigate the risk of both phenomena.

## References

- [1] C. Guzman et al., *Deserción estudiantil en la educación superior colombiana. Metodología de seguimiento, diagnóstico y elementos para su prevención*, Bogotá: Ministerio de Educación Nacional, 2009, p. 156. Available: [http://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-254702\\_libro\\_desercion.pdf](http://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_desercion.pdf)
- [2] MinEducación, “Pruebas Saber”, *Mineduación*, Aug. 26, 2010. Available: <http://www.mineduacion.gov.co/1759/w3-article-244735.html>
- [3] Centro de Estudios sobre Desarrollo Económico, CEDE, “Informe Determinantes de la deserción. Informe mensual sobre el soporte técnico y avance del contrato para garantizar la alimentación, consolidación, validación y uso de la información del SPADIES”, Centro de Estudios sobre Desarrollo Económico, CEDE, Facultad de Economía, Universidad de los Andes, Bogotá, Colombia, Feb. 14, 2014. Available: [http://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-254702\\_Informe\\_determinantes\\_desercion.pdf](http://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-254702_Informe_determinantes_desercion.pdf)
- [4] N. Suárez-Montes & L. B. Díaz-Subieta, “Estrés académico, deserción y estrategias de retención de estudiantes en la educación superior”, *Revista de Salud Pública*, vol. 17, no. 2, pp. 300-313, 2015. doi: 10.15446/rsap.v17n2.52891
- [5] M. Astudillo, “Factores de riesgo asociados a la deserción estudiantil en la Universidad ICESI”, Tesis de grado, Facultad de Ciencias Administrativas y Económicas, Programa de Economía, Universidad ICESI, Santiago de Cali, Colombia, Mayo, 2013. Available: [https://repository.icesi.edu.co/biblioteca\\_digital/bitstream/10906/71417/1/factores\\_riesgo\\_asociados.pdf](https://repository.icesi.edu.co/biblioteca_digital/bitstream/10906/71417/1/factores_riesgo_asociados.pdf)
- [6] J. C. de la Rosa, R. Linero, I. Meriño & O. Rodríguez, “Acompañamiento de estudiantes en alto riesgo académico”, in *Encuentro Internacional de Educación en Ingeniería ACOFI 2014*, Cartagena de Indias, Colombia, Oct. 7-10, 2014, pp. 118-119.
- [7] F. Barrero, C. Barrero, H. Borja & M. Montaña. “Factores de riesgo asociados a la deserción estudiantil universitaria en programas de pregrado de la Universidad de San Buenaventura, Bogotá (2009-2013)”, *Academia y Virtualidad*, vol. 2, no. 8, pp. 60-72, 2015. Available: <http://revistas.unimilitar.edu.co/index.php/ravi/article/view/1423>. doi: //dx.doi.org/10.18359/ravi.1423
- [8] E. P. Ibarra García & P. Medina Mora, “Model prediction of academic performance for first year students”, in *MICAI 2011: 10th Mexican International Conference on Artificial Intelligence*, Puebla, Mexico, Nov. 26-Dec. 4, 2011, pp. 169-174.
- [9] S. Taruna & M. Pandey, “An empirical analysis of classification techniques for predicting academic performance”, in *2014 IEEE International Advance Computing Conference (IACC)*, Gurgaon, India, Feb. 21-22, 2014. Available: <http://ieeexplore.ieee.org/document/6779379>. doi: 10.1109/IA-DCC.2014.6779379
- [10] G. Gray, C. McGuinness & P. Owende, “An application of classification models to predict learner progression in tertiary education”, in *2014 IEEE International Advance Computing Conference (IACC)*, Gurgaon, India, Feb. 21-22, 2014. Available: <http://ieeexplore.ieee.org/document/6779384>. doi: 10.1109/IA-DCC.2014.6779384
- [11] N. Thai Nghe, P. Janecek & P. Haddawy, “A comparative analysis of techniques for predicting academic performance”, in *37th Annual Frontiers in Education Conference. Global Engineering: Knowledge Without Borders - Opportunities Without Passports*, Milwaukee, Wisconsin, Oct. 10-13, 2007. Available: <http://ieeexplore.ieee.org/document/4417993>. doi: 10.1109/FIE.2007.4417993
- [12] S. M. Merchán Rubiano & J. A. Duarte García. “Formulation of a predictive model for academic performance based on students’ academic and demographic data”, in *2015 IEEE Frontiers in Education Conference (FIE)*, El Paso, Texas, Oct. 21-24, 2015. Available: <http://ieeexplore.ieee.org/document/7344047>. doi: 10.1109/FIE.2015.7344047.
- [13] S. M. Merchán Rubiano & J. A. Duarte García. “Analysis of data mining techniques for constructing a predictive model for academic performance”, *Latin America Transactions*, vol. 14, no. 6, pp. 2783-2788, 2016. Available: <http://ieeexplore.ieee.org/document/7555255>. doi: 10.1109/TLA.2016.7555255.

- [14] S. M. Merchán Rubiano & J. A. Duarte García, "Analysis of data mining techniques for constructing a predictive model for academic performance", in *UNESCO-UNIR ICT & Education Latam Congress 2016*, Online, June, 22-24, 2016, pp. 39-48.
- [15] Machine Learning Group at the University of Waikato, *Weka 3: Data Mining Software in Java*, The University of Waikato, Hamilton, New Zealand. Available: <http://www.cs.waikato.ac.nz/ml/weka>