Prediction of University Dropout Using Machine Learning

Chapter · January 2021

DOI: 10.1007/978-3-030-68285-9_38

3 authors:

Aracelly Nuñez-Naranjo
Universidad Tecnológica Indoamérica

Manuel Ignacio Ayala Chauvin
Universidad Tecnológica Indoamérica

Genís Riba Sanmartí
Universitat Politècnica de Catalunya

Prediction of University Dropout Using Machine Learning

Aracelly Fernanda Núñez-Naranjo[1], Manuel Ayala-Chauvin[2(✉)], and Genís Riba-Sanmartí[3]

[1] Universidad Facultad de Humanidades y Artes, Universidad Nacional de Rosario, Rosario, Argentina
arafer@hotmail.es
[2] SISAu Research Group, Facultad de Ingeniería y Tecnologías de la Información y Comunicación, Universidad Tecnológica Indoamérica, Ambato, Ecuador
mayala@uti.edu.ec
[3] Centro de Diseño de Equipos Industriales, Universidad Politécnica de Cataluña, Barcelona, Spain
genis.riba@upc.edu

Abstract. University dropout is a complex issue that affects all higher education institutions worldwide. This phenomenon is shown by the high proportion of students that never finish their university training, with the associated economic and social costs. The challenge for higher education institutions is to design and improve policies to increase student retention, specially within the first years. This study uses data mining to find patterns and student clustering that help explaining university dropout. The data for the analysis was gathered from the students that signed up on two admission perioof the Universidad Tecnológica Indoamerica of Ambato, Ecuador. A k-means algorithm is used to classify and define the performance patterns, and predictions for new students are made using a support-vector machine (SVM) model. The results allow institutions and the faculty to focus in high risk groups during the first terms and amend their future learning behaviour. To sum up, this study presents a models to explain and predict university dropout, and to design actions to reduce it.

Keywords: University dropout · K-Means · Support vector machine

1 Introduction

University dropout is a phenomenon occurring in higher education institutions all over the world. Due to this ubiquity, it has been studied from multiple theoretical and methodological angles, as well as analysed from both social and economic perspectives [1–3]. University dropout constitutes a social issue [4–7] that affects not only the students themselves, but also their families and governments [8], since training a university student all the way to graduation entails a significant amount of resources (economic and otherwise), and dropping out in effect wastes most of these resources.

Á. Rocha et al. (Eds.): ICITS 2021, AISC 1330, pp. 396–406, 2021.
https://doi.org/10.1007/978-3-030-68285-9_38

This can lead to social inequality and increased poverty [9, 10]. Furthermore, it affects the learning process of university students [11].

University dropout is defined as the abandonment, prolongation or interruption of the university training [12] due to an intentional or unintentional personal circumstance [13] in which the student decides to abandon their studies [14].

Developed countries such as Germany or Switzerland had a dropout rate between 20% to 25% and 7% to 30% respectively in 2006 [15], although there is a high variability even among richer countries. For instance, Croatia had a 4.3% desertion rate in 2011, while Italy and the USA reached 55%.

Dropout rates directly affect graduation rates, which range from as low as 30% up to above 90% [16].

Latin America featured high desertion rates in 2017, around 47%, although with large differences from country to country. For instance, Dominican Republic had a desertion rate of 76%, followed by countries such as Bolivia (73.3%), Uruguay (72%), Brazil (59%), Chile (53.7%), Mexico (53%), Venezuela (52%), Honduras (40%), Argentina (49%), and Colombia (40%) [17]. At the other end of the spectrum, Cuba only had a 25% desertion rate thanks to the implementation of education policies such as tutoring, collaborative learning, learning communities and academic support, all of which have increasing retention as a main goal [18, 19].

The socioeconomic environment is one of the main drivers of dropout. Factors such as marital status, father's education level, career counselling, family wealth, or academic performance have proved to be of interest when studying this issue [20–23]. Furthermore, the widespread poverty and pervasive conflict situation prevalent in many Latin American countries adds up to the many social, academic, and economic variables that increase university dropout.

In Ecuador, university dropout rate was 26% in 2017. This situation can be considered a risk factor for families, society and the higher education network as a whole [24]. To address that, universities and government agencies are making an effort to improve retention and graduation rates through policies and regulations [25].

Assessing university dropout is a very complex problem, with many variables and factors that may contribute in different capacities in different contexts, and that are often difficult to quantify or evaluate, such as adaptation to the academy, motivation, age, gender, sex, parents' education, sociocultural background, or place of origin, among many others [26].

Most dropouts occur during the first two of university life, which may be caused by the onset of new goals or life projects and influenced by the personal, familiar and social factors of each student. On the other hand, motivated students with realistic expectations, clear goals, and acceptable academic performance do not abandon their training and have the highest chance to successfully achieve graduation [27].

Ramirez-Rueda et al. characterized the dropout risk as an integration of variables related to the socioeconomic, institutional, academic and personal status. Likewise, Moncada-Mora created a mathematical model of the dropout risk that integrates academic and non-academic variables, such as age, gender or housing. This model estimates the likelihood of a student to drop out, or otherwise stay in their university training. This study also determined that the decision to abandon the studies is often driven by factors related to the inadequate interaction between students and faculty [28].

In the light of the above, it is revealed that there is a lack of academic integration between student body and faculty due to the lack of vocation of the professors [29], and the lack of career counselling offered to the students, together with a lacklustre education beginning as early as middle school [30–32]. All of this affects the quality, pertinence and efficiency of university training [33].

One approach to analyse higher education dropout is the use of data mining techniques. These techniques enable multivariate studies, where data sets are grouped in personal data, academic data, and other data generated during the school life. The variables used in the different studies were entry age, sex, passed courses, quantity, and origin. The analysis of these data yielded behavioural patterns that can help predicting dropouts, and therefore use retention strategies to better understand this phenomenon and reduce its incidence [34–36].

Data mining coupled with artificial intelligence (AI) techniques is a powerful tool to store, process, classify and recognize patterns. Furthermore, it enables the implementation of algorithms such as k-means to classify and explain the characteristics of dropout students [37, 38]. Dropout patterns can thus be determined from the personal, academic and work-related variables [39].

Support Vector Machines (SVM) have been widely used to classify information and predict dropout [40]. However, non-integrated models could generate inaccurate or incorrect outcomes [41]. Therefore, this research proposes an algorithm that integrates k-means, restriction criteria and SVM to explore patterns and clusters in order to explain university dropout.

The experimental data for this study was obtained from two student groups of the Universidad Tecnológica Indoamérica in Ambato, Ecuador. A k-means algorithm was used to classify and define performance patterns. Then, an SVM model is used to predict the behaviour of students in the first terms. Finally, there is an explicative and predictive analysis of the personal, academic and economic variables that have an influence on the students' decision, may it be voluntary or coerced, to abandon university training.

2 Methods

The goal of this study is to define a set of statistical measures to determine the factors that have an effect on university dropout in Ecuador, and more precisely the province of Tungurahua. The k-means method is used to sort and classify the data, and the result of this analysis will be used to improve the student retention ratio during the first year.

The sample universe are the university level students in the Tungurahua province. The sample are the 1,078 students admitted on the first term on the two periods of 2014, on seven different degrees (industrial engineering, general psychology, urban architecture, business administration, digital design and multimedia, accounting and auditing, system engineering) in two modes (blended and on-site learning) on two sites of UTI, a private Ecuadorian university.

The approach used consists of 5 steps. First, data collection. For the study case, data was obtained from the Academic Management System (SGA) database. Second, data processing and statistical normalization to allow for further analysis. Third, generate vectors of the main characteristics, which combine the available data for grades, terms taken, efficiency index, etc. Fourth, group students into behavioural categories with a clustering techniques. Finally, use the resulting data to make predictions using Support Vector Machine techniques.

Each student is described with a vector of quantitative data, including grade average, efficiency index, and number of terms, which are used to assess the academic performance. Equation (1) represents a group of students x:

$$x = (Progen, Propdr, Indefi, Promapro, Semcur, Mention, Econom)^T \quad (1)$$

Where: Progen – general grade average, Propdr – weighted grade average, Indefi – efficiency index, Promapro – average grade of passed subjects, Semcur – number of terms, Mention – career, Econom – economic situation.

Table 1 shows the algorithm created for this research. It is divided in three parts. The first part creates clusters with their centroids, and calculates the statistical values (mean, standard deviation, maximum and minimum values) of the graduates. The second part sorts the graduates based on a quantitative restriction. The third part generates a prediction with the data of the first year students. It includes the prediction functions kernel "lineal" and kernel "poly" from the Python suite.

Table 1. Algorithm to cluster, classify, measure and predict

Algorithm	K-means, statistical measures and prediction
Input:	$\boldsymbol{x} = (Progen, Propdr, Indefi, Promapro, Semcur, Mention, Econom)^T$
Output:	Number of clusters, statistical measures, classification and prediction
	Initialize - K Clusters with their centroids of $\mu_1, \ldots, \mu_k$ randomly
	While not converge:
	for i in range ($dataset$):
	$C_k = argmin\|x_i - \mu_k\|^2$
	for j in range ($\boldsymbol{k}$):
	$\mu_j = \frac{1}{N}\sum_{i=1}^{N} x_i$
	end while
	Return Clusters and statistical measures
	Initialize – classification with restriction criteria
	if $Propdr \leq 10$ and $Indefi \leq 0.7$
	class A
	Else
	$Propdr > 10$ and $Indefi > 0.7$
	class B
	end classification
	Return Table of classification of students using class labels A and B
	Initialize – prediction model SVM
	Model (Kernel = 'linear') - $f(x) = sgn(w^T x + b)$
	Model (Kernel = 'poly') - $k(x_i, y_j) = (x_i \cdot y_j)^d$
	Input Data for prediction
	$\boldsymbol{x}^* = (Progen, Propdr, Indefi, Promapro)^T$
	end prediction
	Return Table of prediction of students using class labels 1 and -1

It is important to point out that measuring academic performance is a very complex and multidimensional problem, which measures in a determined way the benefits of the learning process. It involves the characteristics of the student, of their social environment, and the education system. Therefore, there are many definitions of academic performance, but all of them agree that one indicator are the grades, which quantitatively measure the student's achievements. The efficiency index is another academic performance indicator that is considered in this research.

Finally, supervised classification methods based on linear al polynomic models are used to produce the forecast for the first year students. These methods are used in parallel during the third part of the algorithm in order to compare and validate the results of the prediction

3 Results

This sections analyses the results obtained from the algorithm when applied to the data provided by the UTI. In particular, from 376 students who graduated in 2015. For each student, the file contains the following attributes: general mean grade, weighted mean grade, efficiency index, passed subjects mean grade, and number of terms.

Figure 1 shows the efficiency index (Indefi) compared to the number of terms (Semcur) data in $\mathbb{R}^2$.

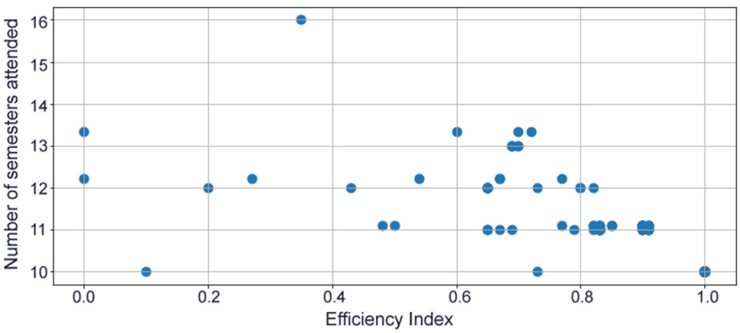

Fig. 1. Data display number of semester's vs efficiency index

3.1 K Clusters

The "elbow" method was used to determine the number of clusters for the analysis. This method allows to select the number of clusters adjusting the model with a range of k values, in order to find a value of k that causes a sudden change in the slope. In this case, k = 7 is a good fit. Then, the K-Means function was used to obtain the best configuration with 7 groups (1,000 iterations).

Figure 2 shows the results of applying the algorithm to the set of data, confronting the number of terms taken with the efficiency index.

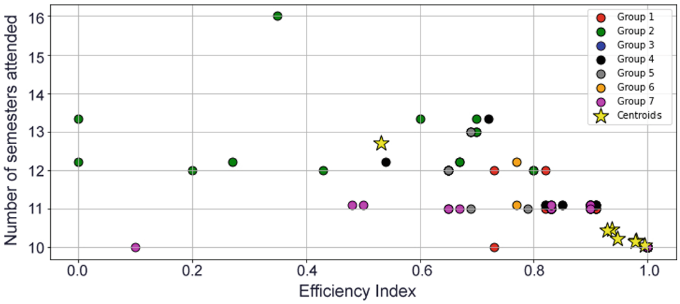

Fig. 2. Clusters with centroids.

Cluster 5 is the largest one, which includes 109 students with a very good grade average of 8.22. Clusters 6 and 7 have 18 and 23 students, with good average grades of 7.86 and 7.92, respectively.

3.2 Classification with Restriction Criteria

The statistical measures show that the weighted grade average and the efficiency index are both good parameters to classify the graduates. The clusters have distinct values of these parameters, which in turn allow for further classification:

Class A, labeled −1, includes all graduates that

$$\text{Propdr} \leq 10 \ \text{ and Indefi} \leq 0.7 \tag{2}$$

Class B, labeled 1, includes all graduates that

$$\text{Propdr} > 10 \ \text{ and Indefi} > 0.7.7 \tag{3}$$

The result of this restricted classification is 29 students in Class A, and 347 students in Class B. Table 2 shows the statistical measures for both Classes A and B.

Table 2. Statistical measures for Classes A and B

Statistical measures	Class A			Class B		
	Propdr	Indefi	Semcur	Propdr	Indefi	Semcur
ME	8.02	0.54	12.22	8.38	0.98	10.19
DE	0.43	0.21	1.09	0.50	0.06	0.46
Min	7.39	0.00	10.00	7.42	0.72	10.00
Max	9.10	0.70	16.00	9.83	1.00	13.33

3.3 SVM Prediction

Finally, once the graduates' data is sorted in two Classes, a Support-vector Machine (SVM) algorithm is used to obtain the prediction. SVM is a classifier formally defined by a hyperplane that separates the groups. In other words, given a set of labeled training data (supervised learning), the algorithm generates an optimal hyperplane that categorizes new students, with each class being on one side of the hyperplane.

The experiment was realized with a sample of 12 students attending the first year. A model was built using the SVM in order to predict which students will have a low performance. The results show that there is a single student in Class A. Therefore, teachers have to take corrective actions and design actions to reduce dropout. Furthermore, there are social, economic, institutional and cultural factors that can have an effect on the behaviour of the students.

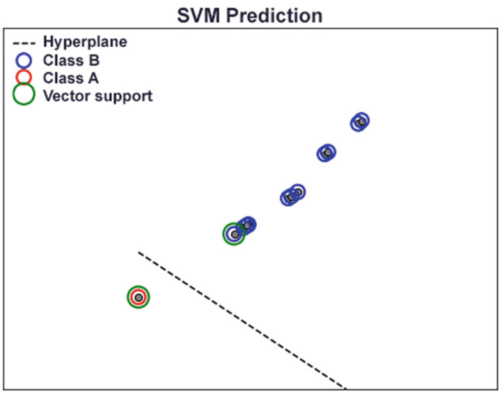

Fig. 3. Support-vector machine prediction.

Figure 3 shows the supervised classification of the sample. It displays the support vector (green), class B (blue) and A (red), and the hyperplane separator, which separates the student predicted to be in Class A.

A linear separation was applied using the Python library $kernel = '\ linear'$. In order to compare results, a second polynomial separation using the Python library $kernel = 'poly'$ was applied too. Both models, lineal and polynomic, produce the same results, that is that one student will have low performance due to economic and social issues during their college years.

4 Discussion

Results shown in previous sections display the power of this method applied to educational data mining. The application developed in Python is flexible and outputs the student performance results both graphically and numerically. With this information, teachers can identify low performance students, improve retention and obtain an overall better performance of the education system.

The models and configurations developed in Python are complex processes that create automatic learning solutions even without programming skills. Therefore, it is replicable by people working in education administration, in research or in learning support tasks, that may need to explore student performance in order to support education institutions.

The Universidad Tecnológica Indoamérica is a private higher education institution. Its dropout rate on 2014 was 44.9%, while its graduation rate in 2019 was 17.25%. This shows an academic lag in the graduation process due to failing and retaking courses, degree swaps, and re-entries. These rates result in a high academic failure rate when compared to similar institutions.

This research focuses on grouping, classifying and predicting using quantitative education data from the Academic Management System (SGA) of the UTI. The SGA was used to acquire data, using 5 data from each admission period in order to predict the students' performance. The total number of students was 376 between February 2014 and July 2015.

The assessment using the K-Means algorithm divides the sample in 7 clusters, which are further classified in classes A and B.

Compared to other algorithms [42, 43], this one requires less pre-processing, since its configuration accepts an input vector that contains all the characteristics of the sample. Therefore, it is very flexible, can be extrapolated, and has a fast processing speed. However, it also has limitations related to the difficulty to code the restrictions to determine student classes, since they are unique for each individual institution. On the other hand, outputs also depend on the purpose of the analysis. Therefore, the code needs to be adjusted for each analysis and educational institution.

Further development of the code should go into blending the prediction algorithm with the SGA of the university, so that student assessment can be obtained in real time, in order to improve retention, and teaching and learning processes.

5 Conclusions

The present research developed a technique to assess and classify the expected academic performance students admitted into the university system based on clustering techniques. The model is based on K-Means to determine the number of clusters, and a classification based on quantitative restrictions for the clustering. The predictions are obtained with supervised learning algorithms using the lineal model $(\text{kernel} = '\text{linear}')$ and the polynomic model $(\text{kernel} =' \text{poly}')$ of the Python suite.

The classification of the student groups based on restrictions improve the interpretability and explainability of the resulting models. Comparison between models produces as a result the value of low performance students in a quick and precise manner. Furthermore, the model is flexible but it also can be generalized. Therefore, we believe that these procedures and tools can be an asset in education data analysis.

The proposed method can be used to estimate the future performance of the students that are admitted to the university system, and thus prepare strategic actions in order to improve retention and graduation rates with learning and support activities according to each student's needs.

On the other hand, it is possible to expand the software so that it can be integrated within the Academic Management Systems to automatically run the analysis every term. It is paramount incorporate characteristics that can help to explain the results and understand why a certain student is more likely to drop out.

To sum up, using machine learning in education allows teachers to assess and analyse data from education environments and, in the end, improve overall academic results.

References

1. González, L.E.: Repitencia y deserción universitaria en América Latina. Inf. sobre la Educ. Super. en América Lat. y el Caribe 2000–2005. **2005**, 156–168 (2005)
2. Diaz, D.B., Garzón, L.P.: Elementos para la comprensión del fenómeno de la deserción universitaria en Colombia. Más allá de las mediciones. Cuad. Latinoam. Adm. **9**(16), 55 (2016). https://doi.org/10.18270/cuaderlam.v9i16.1248
3. Vargas, J.G.C., Rios, L.S.B., Laverde, R.M.: Propuesta Para Aumentar El Nivel Académico, Minimizar La Deserción, Rezago Y Repitencia Universitari. Sci. Tech. **11**(28), 145–150 (2005)
4. Zavala-guirado, M.A.: Factores internos, externos y bilaterales asociados con la deserción en estudiantes universitarios. **4**, 59–69 (2018). https://doi.org/10.24016/2018.v4n1.103
5. Ramírez-Rueda, M., Urrego-Velásquez, D., Páez-Zapata, E., Velásquez, C., Ramírez, E.H.: Perfiles de riesgo de deserción en estudiantes de las sedes de una universidad colombiana. **38** (1), 275–297 (2008) https://doi.org/10.18800/psico.202001.011
6. De Toluca, V.: Factores que inciden en el rendimiento académico de los estudiantes de la Universidad Politécnica del Valle de Toluca Factors affecting the academic performance of students (2017)
7. Tomás, J., Gutiérrez, M.: Aportaciones de la teoría de la autodeterminación a la predicción de la satisfacción académica en estudiantes universitarios Contributions of the self-determination theory in predicting university students' academic satisfaction. **37**, 471–485 (2019)
8. Carvajal-Muquillaza, C.M.: Modelación de la deserción universitaria mediante cadenas de Markov. **34**, 129–146 (2020)
9. Barrios, A.: Deserción universitaria en chile: incidencia del financiamiento y otros factores asociados. **14**(1), 59–72 (2011)
10. González-Ramírez, T., Pedraza-Navarro, I.: Variables sociofamiliares asociadas al abandono de los estudios universitarios. **35**, 365–388 (2017). http://dx.doi.org/10.6018/j/298651
11. Humphrey, K.B.: New sails for the recruitment, retention, and learning ship. About Campus. **13**, 2–3 (2008). https://doi.org/10.1002/abc.239
12. Cabrera, L., Bethencourt, J., Alvarez, P., González, M.: El problema del abandono de los estudios universitarios. **12**, 171–203 (2006)
13. Tinto, V.: Definir la desercion: Una cuestion de perspectiva. Rev. Educ. Super. **18**(71), 160 (1989). https://doi.org/10.1017/CBO9781107415324.004
14. Siegel, M.J.: Reimagining the retention problem : moving our thinking from end-product to by-product. **15**, 8–18 (2011). https://doi.org/10.1002/abc.20043

15. Villamizar-Acevedo, G., Bayona, L.P.: Identification of motivational and sociodemographic factors in deserters psychology estudents at university. **14**, 124–137 (2011)
16. Rico-Parada, A., Suárez-Correa, Y., González-Cárdenas, Y.: Factores relacionados con la permanencia estudiantil en programas de pregrado de una universidad pública 1. **19**, 155–170 (2017). https://doi.org/10.11144/javeriana.ie19-1.frpe
17. Martelo, R.J., Jimenez-Pitre, I., Villabona-Gómez, N.: Determinación de factores para deserción de estudiantes en pregrado a través de las técnicas lluvia de ideas y MICMAC. Espacios. **38**(20) 2017
18. Drake, J.K.: The Role of Academic Advising in Student Retention and Persistence. **16**, 8–12 (2011). https://doi.org/10.1002/abc.20062
19. Tinto, B.V.: Universities as learning organizations. **1**, 2–4 (1997)
20. Corengia, A., Pita, M., Mesurado, B., Centeno, A.: La predicción de rendimiento académico y deserción en estudiantes universitarios. **4827**, 101–112 (2012)
21. Flores-López, O., Gutiérrez-Emmanuelle, L., León-Corredor, O.: Centros de Apoyo y Desarrollo Educativo Profesional para la observación y disminución de la deserción universitaria. **18**(1), 48–62 (2016)
22. Barragán, D., Patiño, L.: Elementos para la comprensión del fenómeno de la deserción universitaria en Colombia. Más allá de las mediciones. **16**, 55–66 (2013)
23. Fonseca, G., García, F.: Permanencia y abandono de estudios en estudiantes universitarios: un análisis desde la teoría organizacional. Rev. la Educ. Super. **45**(179), 25–39 (2016). https://doi.org/10.1016/j.resu.2016.06.004
24. Vivas-Vivas, R.J., Cabanilla-Vasconez, E., Vivas, W.: Relación entre los estilos de aprendizaje y el rendimiento académico del estudiantado de la carrera de Ingeniería Agronómica de la Universidad Central del Ecuador. **43**, 468–482 (2019)
25. León, J.H.: El perfil del tutor de titulación de tercer nivel. caso: Facultad de Ciencias Administrativas de la Universidad de Guayaquil, Ecuador. **6**, 1–12 (2019)
26. Carvajal, C.M., González, J.A.: Variables Sociodemográficas y Académicas Explicativas de la Deserción de Estudiantes en la Facultad de Ciencias Naturales de la Universidad de Playa Ancha (Chile) Sociodemographic and Academic Variables Explaining Student's Dropout in the Faculty of N. **11**(2), 3–12 (2018)
27. Landry, C.C.: Self-efficacy, motivation, and outcome expectation correlates of college students' intention certainty (2003). LSU Doctoral Dissertations, 1254. https://digitalcommons.lsu.edu/gradschool_dissertations/1254
28. Moncada-Mora, L.: La integración académica de los estudiantes universitarios como factor determinante del abandono de corto plazo. Un análisis en el sistema de educación superior a distancia del Ecuador. **17**, 173–196 (2014)
29. Peña-Fernández, M.A.: Los factores pedagógicos influyen en la deserción universitaria Pedagogical factors influence university dropouts. INNOVA Res. J. . **4**(3), 108–115 (2019). https://doi.org/10.33890/innova.v4.n3.2019.996, ISSN:2477–9024
30. Almeida-Lara, L., Cueva, M.C., Romero-Palacios, A.: Análisis de la equidad de género en la Universidad Técnica de Cotopaxi. **VIII**, 259–268 (2017)
31. Rubio, M., Tocain, A., Mantilla, M.: La deserción universitaria en los primeros niveles y la inserción laboral de los graduados. Proyecto Alfa III, DevalSimWeb. Pomtificia Universidad Católica del Ecuador Sede Ibarra. **1**, 26–35 (2012)
32. Baquerizo, R.P., Tam, O.A., López, J.G.: La deserción y la repitencia en las instituciones de educación superior : algunas experiencias investigativas en el Ecuador. **6**, 102–107 (2014)
33. England-Bayrón, C.: Teoría Social Cognitiva y Teoría de Retención de Vincent Tinto: Marco Teórico para el estudio y medición de la auto-eficacia académica en estudiantes universitarios. **5**, 28–49 (2012)

34. Eckert, K.B., Suénaga, R.: Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos. Form. Univ. **8**(5), 3–12 (2015). https://doi.org/10.4067/S0718-50062015000500002
35. Palacios-Pacheco, X., Villegas-Ch, W., Luján-Mora, S.: Application of Data Mining for the Detection of Variables that Cause University Desertion. **1**(2017), 91–103 (2019). https://doi.org/10.1007/978-3-030-05532-5_38
36. Vila: Technology Trends Reading.pdf. **1**(2017), 91–103 (2019). https://doi.org/10.1007/978-3-030-05532-5
37. Vásquez, J., Miranda, J.: Data Science and Digital Business. Springer International Publishing (2019)
38. Treviño, M., Ibarra, S., Castán, J., Laria, J., Guzmán, J.: A framework to avoid scholar desertion using artificial intelligence. Lecturer Notes in Engineering Computer Science, vol. 3 LNECS, pp. 1493–1497 (2013)
39. Fernandez, J., Rojas, A., Daza, G., Gómez, D., Alvarez, A., Orozco, A.: Student Desertion Prediction Using Kernel Relevance Analysis. **1**, 201–209 (2018). https://doi.org/10.1007/978-3-030-01132-1_30
40. Mayra, A., Mauricio, D.: Factors to predict dropout at the universities: a case of study in Ecuador. In: IEEE Global Engineering Education Conference (EDUCON), vol. 2018, pp. 1238–1242, April 2018. https://doi.org/10.1109/educon.2018.8363371
41. Hegde, V., Prageeth, P.P.: Higher education student dropout prediction and analysis through educational data mining. In: 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 694–699 (2018). https://doi.org/10.1109/icisc.2018.8398887
42. Manrique, R., Nunes, B.P., Marino, O., Casanova, M.A., Nurmikko-Fuller, T.: An Analysis of Student Representation, Representative Features and Classification Algorithms to Predict Degree Dropout. In: Proceedings of the 9th International Conference on Learning Analytics & Knowledge, pp. 401–410 (2019). https://doi.org/10.1145/3303772.3303800