International Journal of Mechanical and Production Engineering Research and Development (IJMPERD)
ISSN (P): 2249–6890; ISSN (E): 2249–8001
Vol. 10, Issue 3, Jun 2020, 16075-16086

ACADEMIC INTERRUPTION MODEL USING AUTOMATIC LEARNING ALGORITHMS

LEONARDO E. CONTRERAS [1], HÉCTOR J. FUENTES [2] & Y. JOSÉ I. RODRÍGUEZ [3]

[1]*Faculty professor, Engineering School. Engineering Doctorate student, Universidad Distrital Francisco José de Caldas, Colombia*

[2]*Faculty professor, Engineering School. Doctorate in Economics student, Universidad Distrital Francisco José de Caldas, Colombia*

[3]*Faculty professor, Engineering School. Doctor of Engineering, Universidad Distrital Francisco José de Caldas, Colombia*

ABSTRACT

The definition of student dropout is still under discussion. However, according to researchers, it could be understood as a state of abandonment that can be explained by categories related to different variables: socioeconomic, individual, institutional and academic. The purpose of this article was aimed at proposing a Machine Learning-based model using classification algorithms, in order to establish the capacity to predict the conditions that lead an industrial Engineering student to interrupt or continue his studies. Classification models are built through methodologies such as KNN, SVC, Perceptron and decision trees. The result showed that the seven variables that have the most influence in academic dropout of students in universities are: cohort, approved courses, flunked courses, gender, ICFES score (Colombian state exam) for math aptitude, ICFES score for math condition.

KEYWORDS: Data Analysis, Educational data Mining, Engineering Education, Machine Learning & Model

Received: Jun 08, 2020; **Accepted**: Jun 29, 2020; **Published**: Oct 14, 2020; **Paper Id.**: IJMPERDJUN20201525

Original Article

1. INTRODUCTION

Nowadays, significant changes are taking place in different fields. The educational field is being marked by the implementation of systems used for academic tests, information gathering and data analysis. This is driven by the fact that the institutions have large databases and storage methods, yet these are rarely used. Thus, teachers are forced to adopt new tools and strategies that allow them to identify at-risk students and design new ways to support their learning process (Tempelaar, 2015). This scenario is fairly similar in the technological sector, where large amounts of data from all over the world can be visualized. These data which comes from human – machine interaction are stored in computers and servers from large companies that can suggest specific actions such as recommending articles for consumption. The information analysis is not only useful for students and teachers but it may also facilitate decision-making within the administrative areas of institutions and governments leading the educational sector (García-Tinisaray, Manuel, & Pichardo, 2019). In higher education, the efficiency of the educational process can be measured through specific metrics such as dropout rate, graduation rate and retention rate. This work involves different variables that determine whether a student concludes his undergraduate studies, or whether he is forced to abandon the program due to different reasons (Zambrano, Albarran, & Salcedo, 2018). One of the most important factors in the efficiency of the educational process is the retention parameter, defined as the difference between the number of students that enroll in first semester and those who graduate each year

(Salcedo, 2010). Furthermore, academic performance is the main indicator of the success or failure of students. The factors that have an impact on the possibility of dropout (interruption of studies) of a given higher education program are multivariate (Estrada & Quintero, 2015). These multidimensional aspect causes that the reason for a student to dropout depends on many other factors such as the goals set by teachers, students and institutions (Khan & Choi, 2014), among others.

Hence, data analytics is an emerging field in Colombian education. The use and interpretation of the data generated by students in their academic journey has not been developed enough to determine the influence over the academic performance, dropout and graduation (García-González, Sánchez-Sánchez, Orozco, & Obredor, 2019). In fact, Colombia has evidenced significant growth in dropout rates in higher education programs (Guzmán & Durán, 2009). This problem has prompted the study of the matter in different universities which have captured the academic information contained in the SPADIES open data system to establish the reasons behind student dropout. Given this perspective, the prediction of academic dropout becomes a challenging task. The purpose of this article was aimed at proposing a Machine Learning-based model using classification algorithms, in order to establish the capacity to predict the conditions that lead an industrial Engineering student to interrupt or continue his studies.

1.1 Context

The definition of student dropout is still under discussion. However, according to researchers, it could be understood as a state of abandonment that can be explained by categories related to different variables: socioeconomic, individual, institutional and academic (Tinto, 1990) states that the study of dropout in higher education is extremely complex, since it implies not only a variety of perspectives, but also different types of abandonment. Additionally, (Giovagnoli & Porto, 2002) state that student dropout can be seen as either voluntary or forced, and takes place when a number of enrolled students does not follow the conventional path set by the academic program. This could be caused by dropping out, taking more time than expected to finish the program or leaving the higher education system. The phenomenon of academic dropout is a notorious problem for governments and higher education institutions nowadays since it has effects on labor-related, social, economic and emotional aspects as well as repercussions in students, their families and the institutions. The Ministry of Education has implemented different measures to tackle this phenomenon such as the generation of the SPADIES (System for the Prevention of Dropout in Higher Education) system that allows each institution to identify and classify its students in terms of dropout risk with certain specific variables. The Engineering Faculty from Universidad Distrital has not been exempt from this situation. With approximately 500 students, statistics have shown that 48,7% of students that enrolled in some program between the academic periods 2009-1 and 2017-3 did not graduate, dropped out of the program or lost their student status (García, 2018).

1.2. Disciplines and Variables

Proposing a model for the solution of the student dropout problem requires an integral view up to a certain point. The data must come different sources and actors of the educational process. Said data is analyzed to deliver relevant and useful information and facilitate decision-making processes. Therefore, there is a need to include different disciplines such as Computer Engineering, Mathematics, Machine Learning, Databases and Education (Castrillón, Sarache, & Ruiz, 2020) (figure 1).

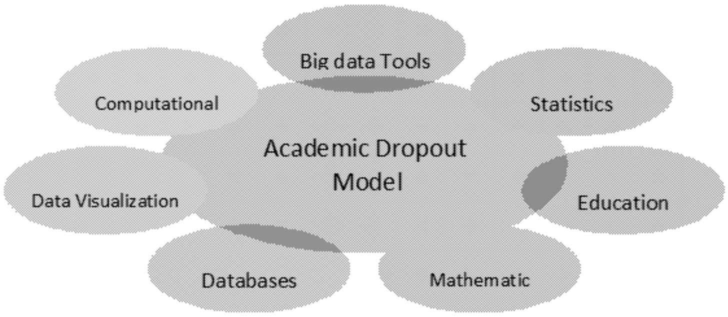

Figure 1: Interdisciplinarity for the Solution of the Problem. Source: Author.

The model intends to predict the status of a student after he has completed 10 academic semesters, he could be an alumnus, retired from the program, abandoned, among others. It also involves different interrelated variables that predict the chances of a student dropping out of the Industrial Engineering program. According to Dussan & Montoya (2016), (Estrada & Quintero, 2015), (Tomás, Expósito, & Sempere, 2014), the factors that could have an influence over his choice are shown in table 1.

Table 1: Variables that Affect Academic Dropout. Source: Author

Before starting at the university	Academic factors	Global ICFES score, ICFES Score in Math, ICFES Score in Physics, ICFES Score in Chemistry, ICFES Score in Spanish, ICFES Score in Social Sciences, ICFES Score in Philosophy, ICFES Score in Natural Sciences, ICFES Score in English.
	Demographic factors	Age, gender, civil status, type of school (private / public) attended during High School, time elapsed before enrolling into university
	Socioeconomic factors	Type of application, place of birth, socioeconomic strata, family income, economic dependence, currently employed
After enrolling at the university	Personal factors	Whether the student has had health issues, problems adapting to the city, change of civil status, calamity at home, conflicts with teachers or other students, how often does he drink alcohol
	Subjects	GPA, number of credits earned, number of credits lost, cohort, average grade in the subjects, number of tests, flunked courses, approved courses.
	Institutional factors	Schedule-related issues and cancellation of courses due to reasons outside of his control, academic interruptions, whether university met his expectations.

1.3 Automatic Learning

The term Big Data has gained popularity partly due to statistical methods that were initially used for analyze data and. The progress attained in Computer Sciences and the rise of Artificial Intelligence (AI) have led to the development of three fields: Robotics (human motion), Natural Language Processing (text and speech) and Machine Learning (data) (Cioffi, Travaglioni, Piscitelli, Petrillo, & De Felice, 2020). Nowadays, Machine Learning (figure 2) has impacted the educational field and can be defined as the science that uses algorithms to make interesting statements based on datasets, without the need to write a code in order to solve a problem (Contreras & Rodriguez, 2018). The most commonly used methods for classification are supervised learning and unsupervised learning.

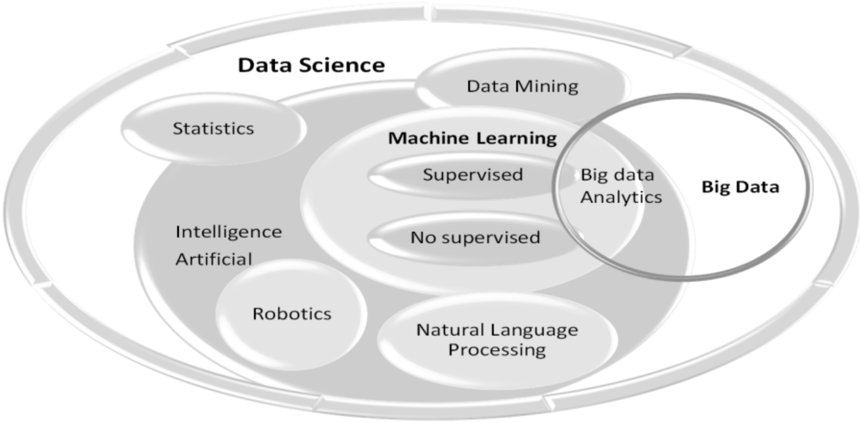

Figure 2: **Data Analysis Overview. Source: Author.**

2. METHODOLOGY

Proposing a methodology to solve the problem can be as complex as the fact of determining which variables could affect the student's resolve. The methodology used is briefly summarized in eight steps: (1) Definition of the population of the study, (2) Variable identification, (3) Loading and cleaning the data, (4) Descriptive statistics, (5) Data transformation, (6) Identification of the main attributes that affect academic interruption, (7) Prediction of academic dropout, (8) Algorithm performance metrics.

Step 1: Definition of the population of the study. Industrial Engineering is one of the programs with the highest number of students enrolled. The analysis involves a total of 3300 records from students over the 2003-2019 period that helped predict the reasons that lead a student to drop out from a program.

Step 2: Variable identification. 30 independent variables were considered in this study that correspond to academic, demographic, course-related and socioeconomic factors as shown in Table 1. For the prediction of the response variable, students could be classified into six causes of academic dropout. Alumni (student who successfully completes the program), excluded (student that is no longer an active student given that university regulations are not met), failed to pass academic probation (the student loses a subject more than 3 times), abandonment (student that does not enroll in two consecutive semesters), dropped out (student who chooses to stop studying), canceled (student who canceled the current semester and never continued his studies) and unofficial enrollment.

Step 3: Loading and cleaning the data. Initially, there were many CSV files with information related to the variables of analysis. They had to be combined to create a data system for each record (student). Therefore, various CSV files were imported using Microsoft Access (database management system) and converted into Access spreadsheets in order to build the input data of the models (Valencia Cárdenas, Correa Morales, & Díaz-Serna, 2015). Afterwards, the data had to be cleaned which led to 3020 records.

Step 4: Descriptive statistics. The initial step of a project involving Machine Learning algorithms is to understand the database through individual charts of the variables and other charts that plot the relationship between independent variables and between said variables and the output variable (Fernandes et al., 2019). Diverse statistics were obtained from the variables mentioned in step 2. This enables to determine how many students drop out of the university in terms of gender, academic period, social strata, etc. It also allows to establish the most frequent reasons of academic dropout. A plot in figure 3 shows the reasons behind academic dropout (target variable) of students over the analyzed period.

Step 5: Transformation of data. Many machine learning algorithms have a better performance when they have a relatively similar scale and/or have a normal distribution (Jahangiri & Rakha, 2015). Hence, the process known as Data transformation is carried out in Machine Learning. Normalization is one type of transformation which is a useful technique to transform attributes from a non-Gaussian distribution with different means and standard deviations. It operates on rows and not over columns as other transformation methods (re-scale and standardize). The normalization of all features makes that all values remain between -1 and 1 (unitary norm or vector with length of 1 in Linear Algebra) (Li & de Rijke, 2019).

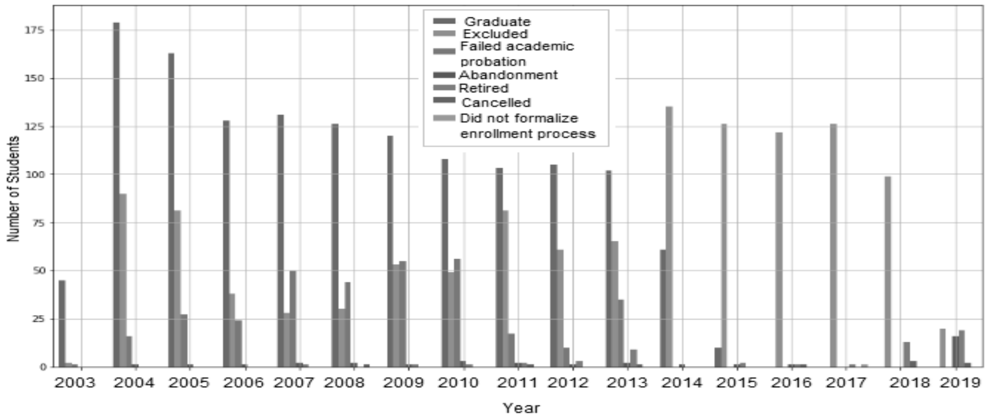

Figure 3: Frequency of Reasons Behind Academic Dropout VS Academic Period.
Source: Calculations made by author.

Step 6: Identification of the main attributes that influence academic dropout. It is necessary to establish whether all variables [25] must be entered into the models or some should be considered to be more influential in the target variable since there are some independent variables (predictors) that may be correlated and affect the reliability of the model. The feature selection methods can help to identify and eliminate unnecessary attributes that do not contribute to the accuracy of a predictive model or even diminish it (Zaffar, Hashmani, Savita, Sajjad, & Rizvi, 2018).

Step 7: Prediction of the academic interruption. Different algorithms were used in the classification task. Some of them are briefly described: I) Support vector machines (SVM) are an effective method of automatic learning that achieves a high accuracy in prediction by understanding the optimal hyperplane of the training set, thus simplifying the issues in classification and regression (Liu, Wang, Wang, Lv, & Konan, 2017); ii) The decision tree is considered one of the most popular approaches to represent classificators in automatic learning. An inverted tree is created by splitting the data into two sets, keeping in mind the value of the most significant differentiator among all input variables. From each node, additional nodes are created that are ramified into other possibilities with metaheuristics (Radhwan A., Abbas A, & Ali S. A., 2017); iii) The perceptron is an artificial neural network (ANN) formed by multiple layers where the first one is the neuron layer that reads the input values, adds all the input variables according to their weights and the partial result is introduced into an activation function that generates the end result. This algorithm is considered to be a robust technique in terms of data manipulation, adaptability and generalization Sánchez & García, 2017); iv) K-nearest neighbors (KNN) is an algorithm based on instances (algorithm that does not explicitly learn a model) meaning that it can classify values by seeking the most similar points of data, learned during the training phase with some type of distance such as the Euclidean, Manhattan or Minkowski (Zhu, Hodgkinson, & Wang, 2018).

Step 8: Algorithm assessment metrics. In order to determine how well the classification algorithms made predictions, different metrics are proposed. These metrics include: (1) Accuracy defined as the number of correct predictions over the total of number of predictions. (2) Precision defined as the ratio of positive instances predicted over the total number of positive cases predicted. (3) Sensibility (recall) defined as the ratio of correctly predicted instances over the total number of positive cases. (4) Specificity defined as the number of correct negative predictions over the total number of negative cases and, lastly, (5) the F1-score defined as the average of precision and sensitivity.

3. RESULTS

The methodology described hereby led to a set of normalized data in Step 5. Data transformation is carried out since it is possible for an independent variable to have more influence over the independent variable given that its numeric scale is higher than in other variables. It was then necessary to normalize the dataset seeking to reduce the effects of the influences (figure 4) and improve the results of the algorithms.

Step 6: Identification of the main attributes that have an influence in academic interruption. In this part of the study, different methods of attribute selection were implemented in order to find the relevant characteristics for the models. These methods were Anova, square Chi, Random Forest, recursive feature elimination, backwards elimination, recursive feature elimination with cross validation and Extra Trees (figure 5). Based on the mentioned methods, it was determined that the most relevant features regarding the response variable were: semester, cohort, GPA, number of tests, approved courses, flunked courses, ICFES score, gender, ICFES score in Math condition, ICFES score for Math aptitude, ICFES score in Physics,. The results indicate that the dependent variable is not the result of a single independent variable and that the analyzed features are not correlated.

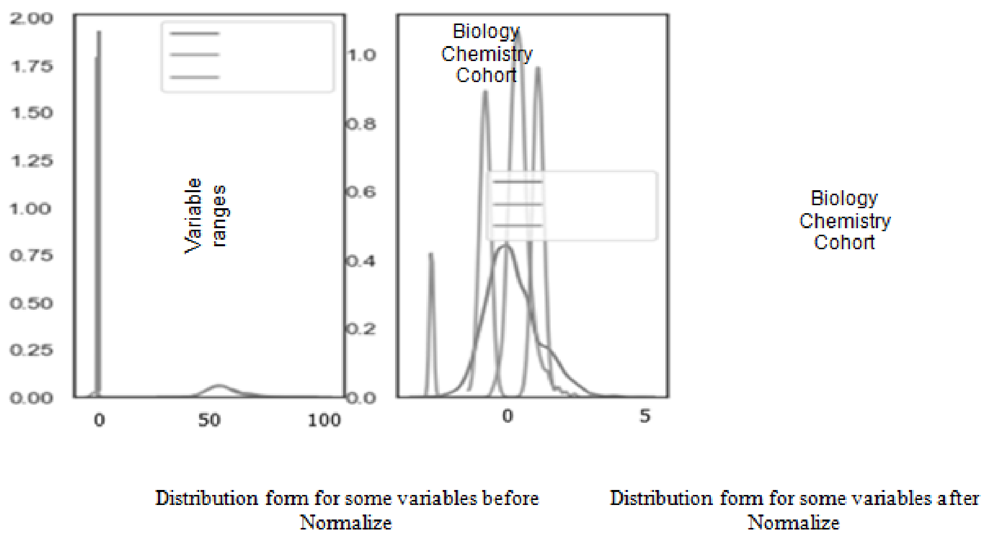

Figure 4: Variable Normalization. Source. Calculations made by authors.

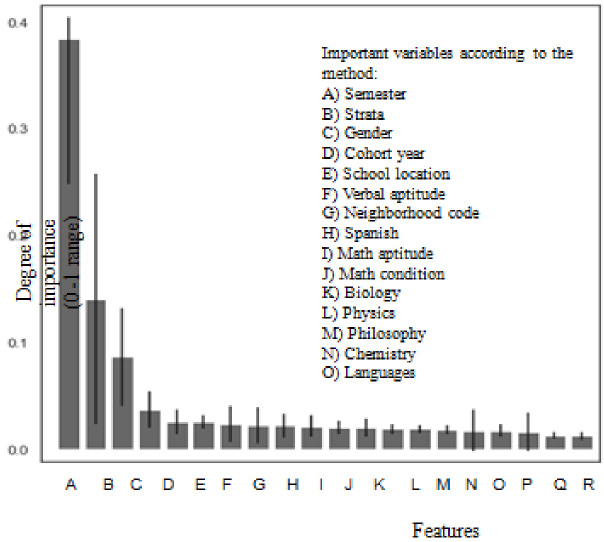

Figure 5: Selection of Variables with the Extra Trees Method.
Source: Calculations made by author.

Step 7: Prediction of academic dropout. Using the selected variables, it was proceeded to separate the data into training data (used to train the algorithms when the response output variable is known) and testing data (number of new data that the algorithm does not know). These data are used to gauge the performance of the algorithm. The training data consists of 906 records and the testing data consists of 2114 records.

Machine Learning algorithms contain parameters that need to be identified and modified in order to obtain better results of the output variable (independent variable). Said parameters are called hyperparameters. The SVM algorithm seeks to generate the best hyperplane (plane in the multidimensional space) that can separate the classes present within a dataset. The hyperparameter Max_Iter is used to iterate until the model reaches the best precision through the algorithm. Figure 6a shows the variation of the model accuracy in face of the variation of the hyperparameter both for the testing data and the training data. The perceptron algorithm is the simplest neural network often preferred in linearly separable data. As seen in other methods, it seeks a separation hyperplane. Figure 6b shows the results for the implemented perceptron algorithm with training and testing data. Another algorithm assessed is the KNN technique which is one of the classical Machine Learning strategies that in this case uses the Euclidean distance. The hyperparameter to assess is n_neighbors (number of nearest neighbors) with a convenient value of 7 or 9 neighbors (figure 6c). Finally, the possibility of building a decision tree was also considered, with the hyperparameter Max_Depth as shown in figure 6d.

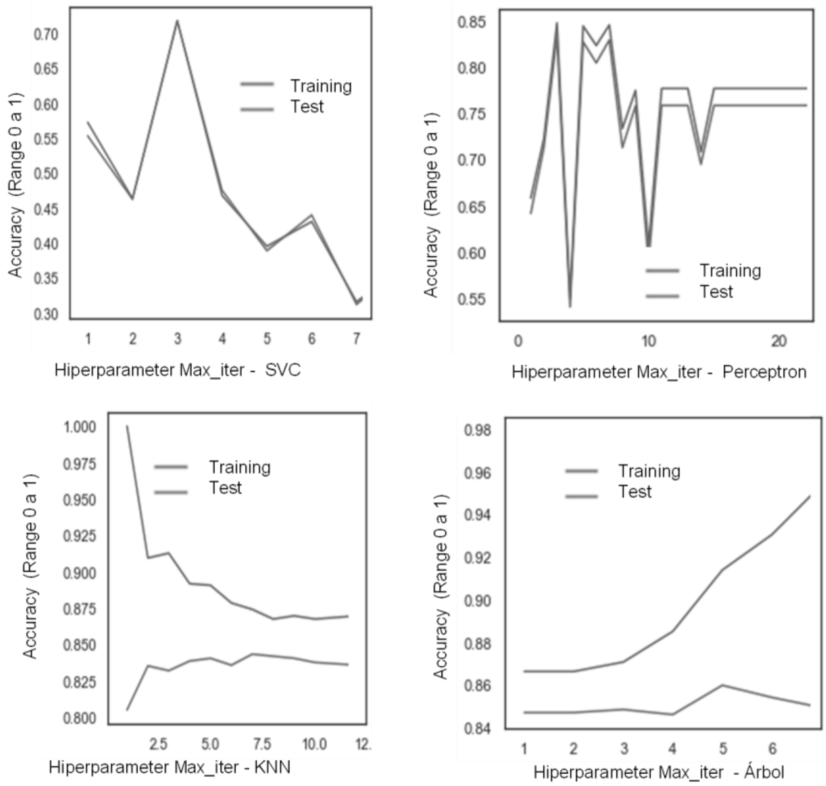

Figure 6: (a) Search of Optimal Max_Iter for SVM
(b.) Search of Optimal Max_iter for the Perceptron Classificator
(c.) Variation of Hyperparameter n_Neighbors in KNN
(d.) Search of Max_Iter for Decision Tree.
Source: Calculations made by autho*r.*

Step 8: Algorithm assessment metrics: It is noteworthy to mention how the algorithms vary in a positive manner until the optimal value is attained for each hyperparameter corresponding to each algorithm. In this case, the variation is shown in Table 2 for the assessment of the accuracy metric. The best hyperparameters for SVM, perceptron, KNN and decision trees are 6, 4, 7 and 7 respectively.

Table 2: Variation of Hyperparameters VS Model Accuracy for Academic Dropout.
Source: Calculations made by *authors.*

Hyperparameter	SVM	Perceptron	KNN	Decision Tree
0	0.805109	0.822668	0.843425	0.847209
2 3 4 5 6	0.832072	0.822668	0.843425	0.848628
4	0.840587	**0.832668**	0.843425	0.855723
6	**0.843425**	0.822668	0.843425	0.855724
7	0.842006	0.812668	**0.843425**	**0.858089**

In terms of the other previously defined metrics, Python has a library called Scikit Learn that includes many functions for the performance analysis in each model. These are accuracy, precision, sensitivity and F1-score. The models show a convenient performance for each metric. In table 3, the accuracy values can be defined as the number of accurate predictions made by the model over the total number of records. If the value is close or equal to 1, it means that all predictions are accurate.

Table 3: Assessment Metrics for the Machine Learning Models.
Source. Calculations made by authors.

Algorithms	Accuracy without Feature Selection Methods	Accuracy with Feature Selection Methods	Precision	Recall	F1
SVC	0.76	0.843425	0.78	0.84	0.80
KNN	0.78	0.843425	0.8	0.84	0.82
Perceptron	0.73	0.832668	0.78	0.74	0.75
Decision tree	0.75	0.858089	0.8	0.84	0.82

4. DISCUSSIONS

The feature selection method allowed to identify the most influential attributes in academic interruption. The work shows that an adequate journey of the student throughout his academic life stems from the variables related with subjects (semester, cohort, GPA, number of tests, approved courses, flunked courses), academic variables (ICFES score, ICFES score in math aptitude, ICFES score in math condition, ICFES score in Physics) and demographic variables (gender). The reference research led to conclude that not only the variables and factors included in Table 1 can be considered. Personal, economic, institutional, emotional, social, demographic and family-related factors can also be reviewed.

In order to increase the metrics of models such as the decision trees and KNN, it is useful to not only identify other variables that contribute to the output variable (academic dropout) but also to identify the most suitable hyperparameters for each algorithm. It is noteworthy to mention that among the group of factors stated in literature there are some hard to control such as the emotional factors. The results evidence that the best accuracy is achieved with the model derived from the decision tree algorithm with 85.8% followed by the SVM and KNN algorithms with 84.34% accuracy. Furthermore, the assessment metrics of the algorithms are improved when using feature selection methods since only the most significant attributes are considered. These results also show that it is possible to generate models through supervised algorithms focused on classification.

The construction of a model that encompasses most of the independent variables, listed in the previous sections, requires considerable time. Colombian universities are often reluctant to deliver data on their students. However, these types of analyses do not actually need personal information which is considered to be private. The fact that many researchers cannot access this information has led most of them to perform studies based merely on descriptive statistics over short periods of time while they aim to determine the causes of dropout and low performance of students. Machine learning and data mining are seldom used in this scenario to generate prediction models that can improve the academic conditions of educators.

5. CONCLUSIONS

- According to the work presented and the results obtained, the main conclusions can be stated. The seven variables that have the most influence in academic dropout of students in universities are: cohort, approved courses, flunked courses, gender, ICFES score for math aptitude, ICFES score for math condition.
- Proposing a model that can predict the reasons behind academic dropout in Industrial Engineering is the first step. This is relevant on a social and economic scale not only for students but also for the universities given that they

can make assertive decisions based on data. It also offers benefits for students at risk that need help from the university and move on with their academic life.

REFERENCES

1- *Castrillón, O., Sarache, W., & Ruiz, S. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. Revista Formación Universitaria, 13(1), 93–102. https://doi.org/10.4067/S0718-50062020000100093*

2- *Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., & De Felice, F. (2020). Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions. Sustainability, 12(2), 502. https://doi.org/10.3390/su12020492*

3- *Contreras, L., & Rodriguez, J. (2018). Big data: An exploration toward the improve of the academic performance in higher education. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 10943 LNCS, pp. 627–637). Springer Verlag. https://doi.org/10.1007/978-3-319-93803-5_59*

4- *Dussan, L., & Montoya, C. (2016). Análisis a partir del Sistema para la prevención de la Deserción de la Educación Superior - SPADIES. Revista Latinoamericana de Estudio Educativos, 12(1), 132–158. Retrieved from http://vip.ucaldas.edu.co/latinoamericana/downloads/Latinoamericana12(1)_8.pdf*

5- *Estrada, J., & Quintero, R. (2015). Bajo rendimiento académico en la Universidad Distrital Francisco José de Caldas. (Universidad Distrital, Ed.) (Primera). Bogotá: Editorial UD.*

6- *Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, G. Van. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. Journal of Business Research, 94(February 2018), 335–343. https://doi.org/10.1016/j.jbusres.2018.02.012*

7- *Garbanzo, & María, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. Revista Educación, 31(1), 43–63. Retrieved from http://www.redalyc.org/articulo.oa?id=44031103*

8- *García-González, J. R., Sánchez-Sánchez, P. A., Orozco, M., & Obredor, S. (2019). Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia Knowledge Capture for the Prediction and Analysis of Results of the Quality Test of Higher Education in Colombia. Revista Formación Universitaria, 12(4), 55–62. https://doi.org/10.4067/S0718-50062019000400055*

9- *García-Tinisaray, D., Manuel, J., & Pichardo, M. (2019). Learning Analytics as an analysis factor of university academic performance. In E. S. P. del L. (ESPOL) (Ed.), CEUR Workshop Proceedings, (pp. 42–50). Retrieved from http://ceur-ws.org/Vol-2231/*

10- *García, R. (2018). Estadística de la permanencia, graduación y deserción de los estudiantes en la Facultad de Ingeniería en programas de pregrado 2009-2017. Bogotá-Colombia: Retrieved from www.udistrital.edu.co*

11- *Giovagnoli, P. I., & Porto, A. (2002). Determinantes de la deserción y graduación universitaria: Una aplicación utilizando modelos de duración 1. Buenos Aires. Retrieved from www.depeco.econo.unlp.edu.ar*

12- *Guzmán, C., & Durán, D. (2009). Deserción estudiantil en la eduacion superior colombiana. Ministerio de Eduacion Nacional- Republica de Colombia, 158. Retrieved from http://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_desercion.pdf*

13- *Jahangiri, A., & Rakha, H. A. (2015). Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data. IEEE Transactions on Intelligent Transportation Systems, 16(5), 2406–2417. https://doi.org/10.1109/TITS.2015.2405759*

14- *Khan, I. A., & Choi, J. T. (2014). An Application of Educational Data Mining (EDM) Technique for Scholarship Prediction. International Journal of Software Engineering and Its Applications, 8(12), 31–42. https://doi.org/10.14257/ijseia.2014.8.12.03*

15- *Li, X., & de Rijke, M. (2019). Characterizing and predicting downloads in academic search. Information Processing and Managenement, 56(1), 394–407. https://doi.org/10.1016/j.ipm.2018.10.019*

16- *Liu, C., Wang, W., Wang, M., Lv, F., & Konan, M. (2017). An efficient instance selection algorithm to reconstruct training set for support vector machine. Knowledge-Based Systems, 116(15), 58–73. https://doi.org/10.1016/j.knosys.2016.10.031*

17- *MEN. (2017, May 6). MEN-Sistema Nacional de Información de la Educación Superior. Retrieved 7 May 2020, from https://www.mineducacion.gov.co/sistemasinfo/Informacion-a-la-mano/212400:Estadisticas*

18- *Radhwan A., Abbas A, & Ali S. A. (2017). Popular Decision Tree Algorithms of Data Mining Techniques: A Review. International Journal of Computer Science and Mobile Computing, 6(6), 133–142. Retrieved from https://www.researchgate.net/publication/317731072_Popular_Decision_Tree_Algorithms_of_Data_Mining_Techniques_A_Review*

19- *Salcedo, A. (2010). Desertion in Colombian Universities. Revista Academia y Virtualidad, 3(1), 50–60. Retrieved from http://www.alfaguia.org/alfaguia/files/1319043663_03.pdf*

20- *Sánchez, P., & García, J. (2017). A new methodology for neural network training ensures error reduction in time series forecasting. Journal of Computer Science, 13(7), 211–217. https://doi.org/10.3844/jcssp.2017.211.217*

21- *Tempelaar, D. T. . et al. (2015). In search for the most informative data for feedback generation: learning analytics in a data-rich context. Computers in Human Behavior, (47), 157–167. https://doi.org/10.1016/j.chb.2014.05.038*

22- *Tinto, V. (1990). Principles of Effective Retention. Journal of the Freshman Year Experience, 2(1), 35–48.*

23- *Tomás, J., Expósito, M., & Sempere, S. (2014). Determinantes del rendimiento académico en los estudiantes de grado. Un estudio en administración y dirección de empresas. Revista de Investigación Educativa, 32(2), 379–392. https://doi.org/dx.doi.org/10.6018/rie.32.2.177581*

24- *Valencia Cárdenas, M., Correa Morales, J. C., & Díaz-Serna, F. J. (2015). Classical and bayesian statistical methods for demand forecasting. A comparative analysis. Revista de La Facultad de Ciencias, 4(1), 52–67. https://doi.org/10.15446/rev.fac.cienc.v4n1.49775*

25- *Zaffar, M., Hashmani, M. A., Savita, K. S., Sajjad, S., & Rizvi, H. (2018). A Study of Feature Selection Algorithms for Predicting Students Academic Performance. International Journal of Advanced Computer Science and Applications, 9(5). Retrieved from www.ijacsa.thesai.org*

26- *Zambrano, C., Albarran, F., & Salcedo, P. A. (2018). Perception of pedagogy students regarding self-regulation of learning. Formacion Universitaria, 11(3), 73–86. https://doi.org/10.4067/S0718-50062018000300073*

27- *Zhu, D., Hodgkinson, L., & Wang, Q. (2018). Academic performance and financial forecasting performance:A survey study. Journal of Behavioral and Experimental Finance, 20, 45–51. https://doi.org/10.1016/j.jbef.2018.07.002*