ISSN2344-8652

Deep Neural Network (DNN) Applied to the Analysis of Student Dropout in a Higher Education Institution

Red Neuronal Profunda (RNP) Aplicada al Análisis de Deserción Estudiantil en una Institución de Educación Superior

Marco Javier Suárez Barón **Juan Sebastian Gonzalez Sanabria**

Jorge Enrique Espindola Diaz

Universidad Pedagógica y Tecnológica de Colombia, UPTC

Recibido: 17/01/2022
Aceptado: 11/04/2022
Publicado: 30/06/2022

Correspondencia de autores:
marco.suarez@uptc.edu.co

Abstract

Objective: This article was built to present the design of a deep neural network (DNN) focused on predicting patterns of student dropout at the UPTC University of Colombia. **Methodology:** We apply specialized artificial intelligence (AI) algorithms for DNN implementation specifically using machine learning for classification and clustering tasks. Additionally, a dataset containing 17 attributes of 3000 academically active students was structured; the data set has been prepared to be trained as an input to the neural network. **Results:** The research analyzes a predictive model trained through the DNN validated by several quality metrics that demonstrate the reliability and accuracy of the results achieved through the model. **Conclusions:** We have focused our study on the application of neural network architecture to address this dropout problem. Our model achieves not only high accuracy, but also a low false negative rate while predicting dropouts on the collected dataset.

Keywords: Learning machine, student dropout, deep neural networks.

Resumen

Objetivo: Presentar el diseño de una red neuronal profunda (DNN) enfocada a predecir patrones de deserción estudiantil en la Universidad UPTC de Colombia. **Metodología:** Aplicamos algoritmos especializados de inteligencia artificial (IA) para la implementación de la DNN específicamente utilizando una máquina de aprendizaje para tareas de clasificación y agrupación. Adicionalmente, se estructuró un dataset que contiene 17 atributos de 3000 estudiantes académicamente activos; el conjunto de datos ha sido preparado para ser entrenado como una entrada a la red neuronal. **Resultados:** En la investigación se analiza un modelo predictivo entrenado a través de la DNN validado por varias métricas de calidad que demuestran la confiabilidad y precisión de los resultados logrados a través del modelo. **Conclusiones:** Hemos centrado nuestro estudio en la aplicación de la arquitectura de redes neuronales para abordar este problema de abandono. Nuestro modelo logra no solo una alta precisión, sino también una baja tasa de falsos negativos mientras predice abandonos en el conjunto de datos recopilados.

Palabras clave: Aprendizaje de máquina, deserción estudiantil, redes neuronales profundas.

Como citar (IEEE): M. Suárez-Barón., J. Gonzalez-Sanabria., y H. Espindola-Diaz. "DRNP Deep Neural Network Applied to the Analysis of Student Dropout in a Higher Education Institution". Investigación e Innovación en Ingenierías, vol. 10, n°1, 202-214, 2022. DOI: https://doi.org/10.17081/invinno.10.1.5607

Introduction

The dropping out of in Higher Education Institutions, (IES- Instituciones de Educación Superior) in Spanish, it is a problematic issue presents across the country. It carried out this problem became one of the most important case studies by control entities in Colombia. For this reason, different universities have been carried out studies looking for explanations, origins and correlations. Most of these studies have been applied usinstatistical techniques which are claimed to find common characteristics of students who drop out their studies. Specifically, in this research, it expects to go beyond of finding some characteristics, it is important to find patterns that can predict when a student is in risk or not of dropping out their studies.

The student who dropping out the university is a relevant topic for all the countries around the world and to find its probable causes is a special subject for this study. For this reason, it has been conducted different studies where has been applied a variety of data analysis techniques which claim to find out which are the main causes to carry out students drop out their academic programs.

To create an algorithm that contains these characteristics, a Learning Machine was used. The final intention is “teach” to the machines to "learn" by themselves without need to be programmed in an explicitly way. In this sense, machines are going to be able to find patterns and generate inference on a set of data. The variables of this research study were conserved in Spanish for getting a better understanding of data analyses.

Background

Detailed below is a synthesis of the literature review carried out around this topic, the student who drop out their studies, where were applied data analytics, data mining and learning machine. Nowadays, the amount of information generated and related to this topic, connect a huge amount of structured data and non-structured too, than can be used to generate value. The application of learning machine or science of data can provide the process, extraction and discovery of useful knowledge, based on this a person or organization can be able to take decisions that benefit in some way [1].

In the case of educational scenarios, both of them, nationally and internationally, different research has been carried out (“Application of data mining techniques for the assessment of academic performance and student drop out of university” [2], “University student desertion and graduation: an application of survival models” [3], “Predictive model of student dropout based on decision trees” [4]. These authors mentioned different techniques to find appropriate factors of dropout having as purpose take decisions who help to decrease this issue.

In India, several studies based on student dropout have been developed using data analysis techniques, for instance, one of them was performed by [5], who through the use of Bayesian classification used models to detect variables that are negatively related.

Specially, the ones which were related to academic performance. After that, having found the relationships, the authors tried to design action strategies to improve this situation.

By the other hand, another study based on student dropout was conducted in India between 2011 and 2013 at a school in Kerala, where considered the information of more than 665 students to carry out data process and classification of them using tools such as WEKA. According to [6], this study helped to assess which are the attributes that predict academic failure. Moreover, among the techniques were used induction rules and decision tree.

In Europe, one of the most meticulous studies was conducted by [7], whom then to do an exhaustive and appropriate research about student dropout in Belgium universities, pointed that the discriminatory analysis, neural network and random forest can be able to get interesting results determining the factors which affect the university students success. Likewise, [8] developed a research doing Top Dow algorithms, Induction Trees (TDIDT) to find factors that interfered with student dropout, based on this research concluded standardized subjects in the first school year was an important factor who determined student dropout.

Another way to look at this situation, according to [9] who conducted a research when used mining techniques, decision trees and Knowledge Discovery in Databases (KDD). As conclusion, determined 84% of student dropout are in strata 2 and they come from South of Colombia municipalities.

In the study developed by [10], applied using neural network, data mining and decision trees whose target was the identification and prediction among the variables related to socioeconomic and academic information obtained across the skill Extract, Transform and Load (ETL).

Finally, [11] conducted a research using different decision trees algorithms as l J48 y el IDE3. The results concluded, after training these algorithms, these are good to predict student dropout, because they had a high similarity of data, all this information is collected in SPADIES1 platform offered by (MEN - Ministerio de Educación de Colombia).

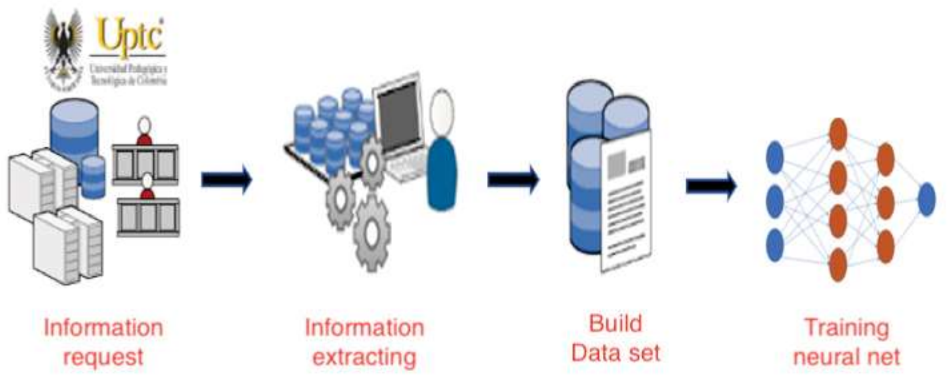
Materials and methods

For the purpose of this research was used a methodology based on four steps. It starts requesting and collecting data from different sources and storage structures. Subsequently, it integrates data from sources as: electronic sheets, online government databases (BD), documents located in institutional repositories whose content is preliminary information corresponds to student permanence and student dropout data obtained in surveys and interviews.

After integrating the information obtained, it is necessary to create a data set called for this research paper "dataset", it is essential and important entry required to apply and train deep neural network (DNN) algorithm designed to analyze the patterns found in the data set obtained. In the Figure 1, presents methodological steps applied in the analyses process of student dropout at UPTC, university. From that methodology applied in the Figure 1, details next the way who was applied the methodology used henceforth refinement and restructuring phase.

1 SPADIES. It is a system that is used as a tool to track the high education students dropout numbers with data provided by High Education Institutions, in Spanish (IES-Instituciones de Educación Superior). This system identifies and weighs behaviors, causes, variables and determined risks to dismiss.

Figure 1. Methodological model used to analyze students dropping out of university



Source: Authors own creation.

On having centralized all the information in a repository was applied cleanliness and data preparation, it was integrated structured and non-structured data, the final “dataset” was confronted; afterward, a data exploratory analysis was applied. Finally, was developed a DNN implementation and training. Once the “dataset” was structured related to the students academic record of UPTC, Sectional Sogamoso, continued to verify the data integrity and analyze different re-structurations to get a “dataset” suitable to DNN training.

Dataset

This step involved a detailed survey of a “dataset” to know and determine what data would benefit neural network training, similarly, project statistics from the data set. Table 1 shows the final structure of dataset obtained after data cleaning.

Techniques of refinement and preparation were necessary to get accurate results of dataset with greater probability of success. In this phase, the tasks are based on empty data cleaning, moreover, the necessary transformations to implement the required algorithms. In addition, the generation of dummies variables is applied for easier implementation of the algorithm. For getting the final data set were applied mapping methods of columns, data deleting and incongruent columns, new columns creation too.

Table 2. Final structure obtained for the data set.

Observaciones	**4216**	
Tipo de archivo	**CSV**	
N° Atributo	Atributo	Estancia
1	id_caso	DISCRETA
2	nombre_facultad	CATEGORICA
3	fecha_nacimiento	CONTINUA
4	ciudad_origen	CATEGORICA

5	ciudad_residencia	NOMINAL
6	sexo	CATEGORICA
7	id_programa	DISCRETA
8	programa	CATEGORICA
9	modalidad	CATEGORICA
10	jornada	CATEGORICA
10	id_estado	DISCRETA
11	nombre_estado	DISCRETA
12	id_periodo_ingreso	DISCRETA
13	descripción_periodo	NOMINAL
14	id_periodo_final	DISCRETA
15	descripcion_periodo	NOMINAL
16	promedio_semestre_final	DECIMAL
17	Promedio	CONTINUA

Source: Own elaboration

Exploratory analysis

In this research paper presents the exploratory analysis from two scenarios. In this sense, Figure 2 shows that most students who have not dropped out maintain a cumulative average of 3.0 or hereinafter, whereas the desertors are a greater amount and maintain averages below of 3.0. Finally, based on the performance demonstrated by DNN, were detected several patterns that visually may affect neural network prediction. By the other hand, the second one scenario, presented in Figure 3, where is possible to evidence obtained dataset shows the direct relationship of students dropout per career. In consequence, the youngest students used to have a high prediction, stating who are most at risk of dropping out their careers than elder students.

Figure 2. Relation between accumulated average vs student drop out status

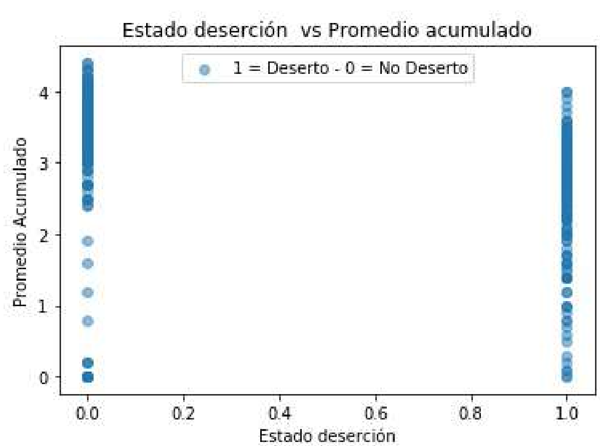
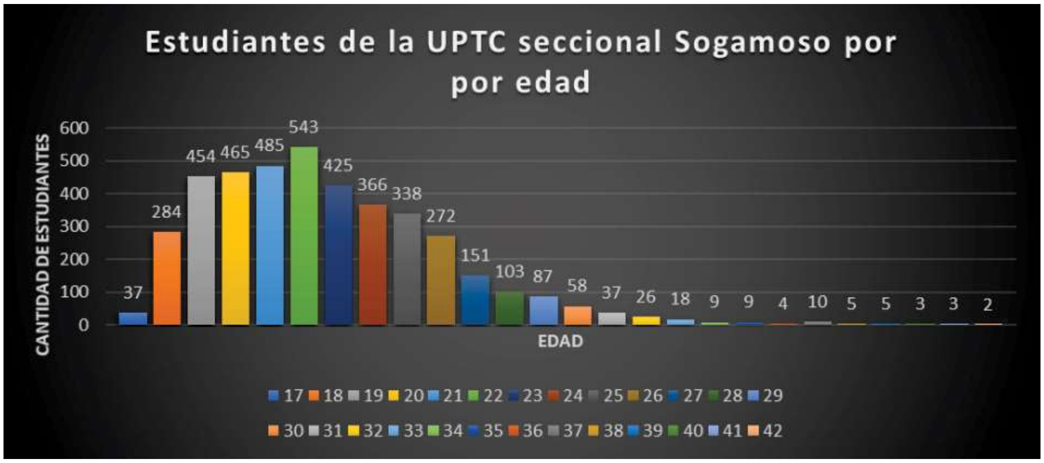


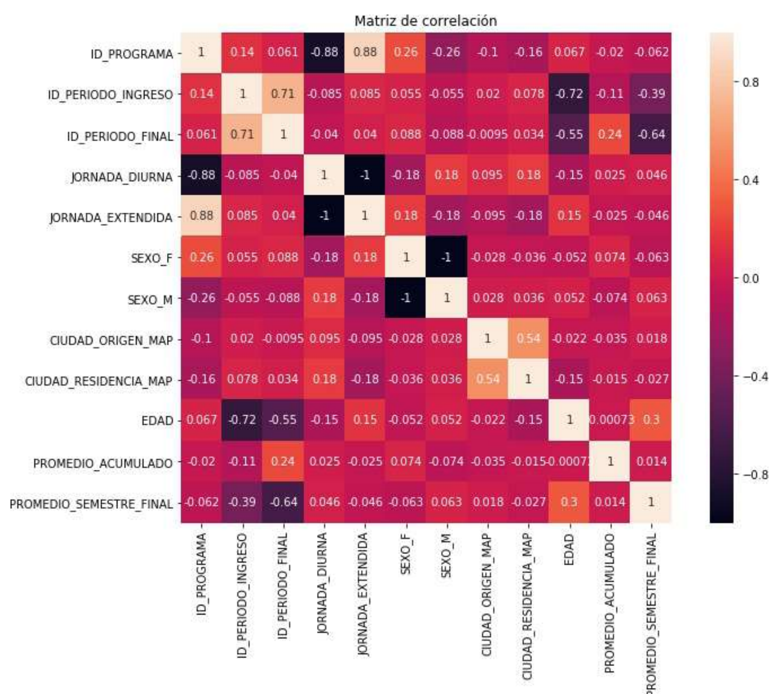
Figure 3. Students number vs ages.



Source: Own elaboration

A way to visualize variable relationships is through correlation matrix who helps to observe how much correlation is among two variables, by using a specific tone. In this case, the matrix results indicates that exists a meaningful correlation among extended journey (JORNADA_EXTENDIDA) and academic program (ID_PROGRAMA) to which the student belongs, in the same way, last lost studied (ID_PERIODO_FINAL) and the period of admission (ID_PERIODO_INGRESO) which are strongly correlated. Finally, appears home city variable (CIUDAD_ORIGEN_MAP) and city of residence (CIUDAD_RESIDENCIA_MAP) also show a high correlation among them, in the Figure 4, is seen the matrix correlation generated.

Figure 4. Correlation matrix among the entrances used by DNN.

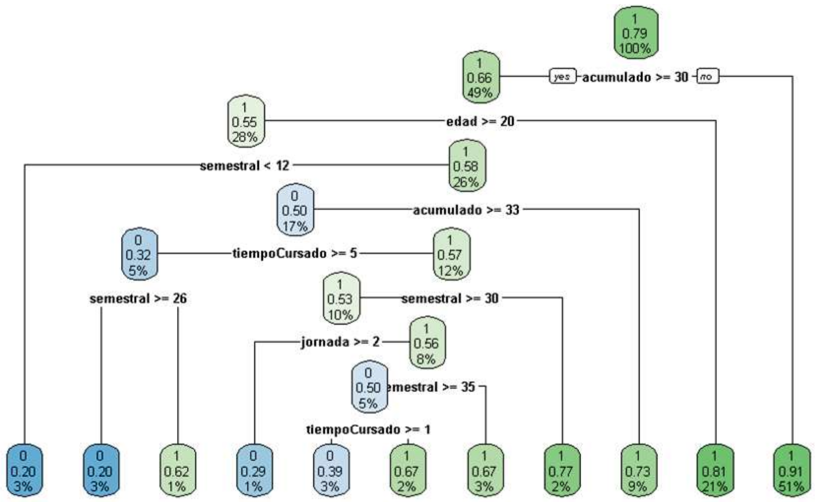


Source: Own elaboration

Results and discussion

For the application and execution of algorithm C.4 the decision tree was taken into account 30% of processed "dataset", to get the following tree illustrated in the Figure 5.

Figure 5. Decision tree executed with test data.



Source: Own elaboration

Table 2 shows decision tree who gets accuracy classification of 80%, an identification of positive elements of 80%, the recall metric gets 98%, it points the most relevant proportion of positive results 80%, in other words, reveals which percent of cases are positive; eventually, the error rate was 21%.

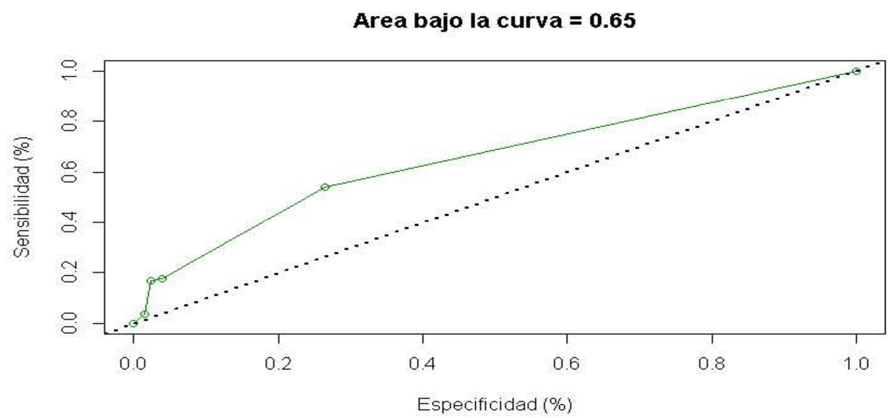
Table 2. Quality metrics of the decision tree

Exactitud	Precisión	Recall	F1_Score	Tasa de Error
0.80	0.80	0.98	0.88	0.21

Source: Own elaboration

To finish test of decision tree the curve generated is "Receiver Operating Characteristic (ROC)" how illustrates Figure 6, obtaining an area under the curve of 0.65, which indicates a good level of prediction based on this methodology, because it offers a higher level of sensivity; that is, catch a highest number of positives cases, in this case, would be a highest prediction level on students who drop out.

Figure 6. Decision tree ROC curve



Source: Own elaboration

For the case of application, the dependent variable is the dropout status to remark these meaningful characteristics within the analysis, where cumulated average of each one student is priority to know if a student is in currently risk to lose its enrollment in the university. In this way, if a student presents averages below of 3.0, it is probably to have a possibility to drop out 49% how shows the first ramification. Lastly, permanence time is going to be a representative variable, because the most of students present a semester average of more than 3.3, who studied more than 4 semesters. In consequence, its dropout was based on academic reasons, whereas the remanding percentage is higher, stayed the enrolment less of 3 semesters, it concludes that the dropping out would be caused by economic and social factors.

In this sense, the process to get the best result have to be conducted through different tests and trials, for this reason, carried out a training process in which the amount of input data varied, the number of neurons used per layer, the number of seasons and the number of observations too. Once trainings were finished, applied a contrast to determine the best assertion result and lowest loss value. In this part of tests took relevance the development of DNN using the multilayer perceptron algorithm. In terms of the model architecture, it was conducted by an activation function for hidden layers[12]. For the output layer was used a sigmoid function. Table 3 states values corresponding to the chosen model of DNN. Regarding to optimization technique, according to [13] was used Adam function, because it has proven to work quite well. The Adam algorithm is an extent of stochastic descent gradient that recently has seen a wider adoption to deep learning applications.

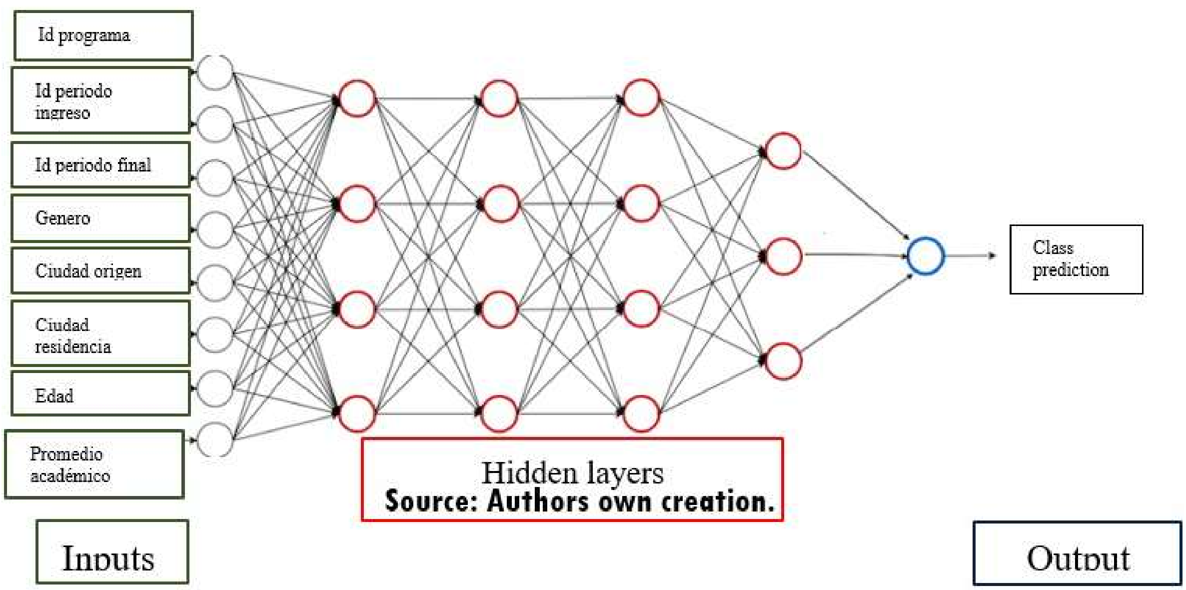
Table 3. Implemented deep neural network model (DNN)

Número de entradas	12
Activación capas ocultas	ReLU
# neuronas capa 1	17
# neuronas capa 2	10
# neuronas capa 3	12
# neuronas capa 4	20
# neuronas capa 5	32
Número de salidas	1

Source: Own elaboration

Figure 7 shows a real representation of deep neural network (DNN) selected and implemented for this research. This DNN contains 12 entries, which were selected a priori in the very beginning and after conducting different trainings[14], it was confirmed that they were the entries that provided the most reliable information from the model, since these entries provided the model with the best predictions. By the other way, to get the number of capable and the number of neurons per layer, it was decided through trial and error until obtaining an acceptable result in model operation, because it did not exist an establish way to know exactly the amount of layers and neurons who will need a model. Basically, the neural network is a black box at the level of prediction interpretation. However, the results can be analyzed by the classificator with specific techniques that allow to recognize how well it works a deep neural network (DNN) among other learning machine algorithms[15].

Figure 7. Neural network model implemented for prediction and training



Source: Own elaboration

The first hidden layer contains 17 interconnected neurons, initially with 12 entries, then 10 neurons constitutes the second hidden layer. In addition, the second hidden layer is interconnected with each neuron who belongs the third hidden layer, it contains 12 neurons, afterwards they interconnect with 20 neurons part of the fourth hidden layer, this layer is interconnecting with neurons in the fifth hidden layer, it contains 32 neurons which are connecting by the exit layer. The exit layer is conformed by only one neuron because the prediction works with a binary class (Deserta=1, No Deserta=0) [16] task that is applicable because just one neuron is used.

Table 3 shows an assessment scenario, where the predictions given by DNN are observed, in this case, when a student belongs systems engineering and computing program, who is a legal age but under 25, has a cumulative average of more than 3 but less than or equal to 4, and a current average of less than 3, on revenue period of the year 2014 onwards. In addition, Table 4(a) illustrates entries got in and the prediction result obtained.

Table 4(a). Prediction results of student dropout generated by DNN

Caso	C. Residencia	C. Origen	Sexo	P. Ingreso	P. Final	Edad	Predicción
1	Duitama	Duitama	M	2016-2	2017-1	23	0
2	Duitama	Tunja	M	2015-1	2017-1	25	0
3	Resto Boyacá	Resto Boyacá	F	2013-1	2016-2	21	0
4	Resto Boyacá	Resto país	F	2013-2	2015-2	22	1
5	Resto Boyacá	Tunja	M	2017-2	2018-1	21	0

Source: Own elaboration

In this scenario was observed that the most student at risk of being a possible deserter according to DNN is a female student who did not born or live nowadays in Sogamoso city and who last semester was studied in 2016. Probably, it points one of the key factors to take into account is how far the university does live the student? [17] how illustrates case number 4. In the second assessment scenario, it was entended to observe predictions thrown by the DNN. By the way, a student who belongs public accounting with cumulative and current average above Boyacá 3.8, where home city and city of residence not is Sogamoso. Table 4(b) shows the entries got in and the prediction result obtained of these entries.

Table 4(b). Prediction results of student dropout generated by DNN

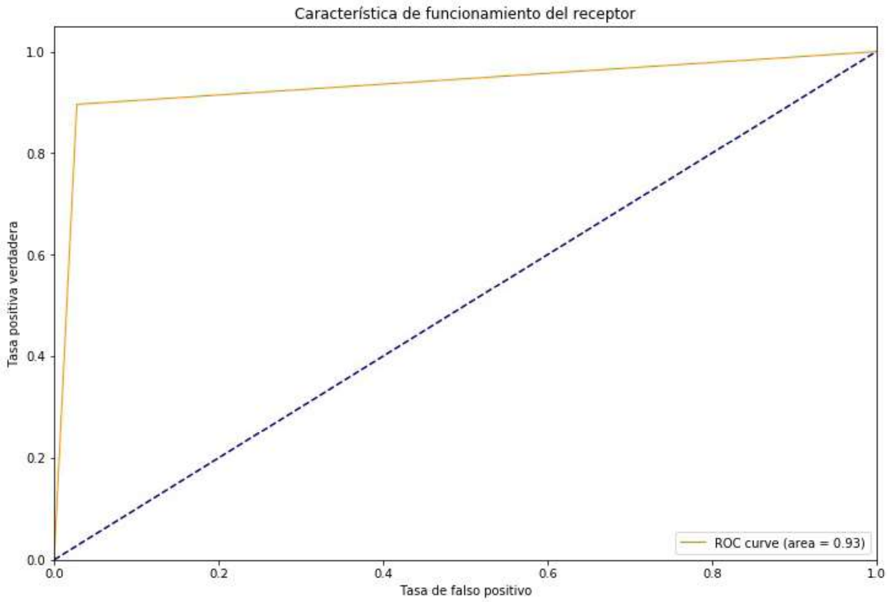
Caso	C. Residencia	C. Origen	Sexo	P. Ingreso	P. Final	Edad	Estado Deserción
1	Sogamoso	Sogamoso	F	2014-1	2017-2	20	0
2	Duitama	Sogamoso	F	2014-1	2018-1	18	0
3	Resto Boyacá	Duitama	M	2014-2	2016-1	23	1
4	Sogamoso	Resto País	F	2017-2	2018-1	21	0
5	Sogamoso	Tunja	M	2016-2	2018-2	22	0

Source: Own elaboration

From the last scenario, it was observed that the most student at risk of being a possible deserter according to DNN, it is a female student, which home city do not belong to Boyacá Department, and city of residence is different to Tunja, Duitama and Sogamoso. This again shows how it looked in scenario 1, that home city and city or residence play an important role when it comes to do a prediction as possible deserter. For instance, in case number three, it can be shown that once again the deserter has a special characteristic: do not live in Sogamoso, the Sectional of UPTC university.

Furthermore, the quality assessment of DNN corresponds to the analysis of "ROC curve" describes in Figure 8. In this research, ROC curve[18] offers an excellent path to interpret how predictive model can distinguish among the truly positive and negative. The model not only requires to correct predictions of the positive as positive, but also the negative as negative. After being implemented this technique to assess the DNN model, it can be said that the classification was performed correctly in 93% of the cases. The ROC curve achieved a high level to 90%, (94% and 93% respectively). It reflects a good accuracy level of the results given by the model. Generally, it points about the constructed model, there is a low probability of getting incorrectly classified values, according to the results provided by the techniques mentioned before.

Figure 8. Applied ROC curve obtained by DNN accuracy analysis



Source: Own elaboration

The data gathered from preparation and cleaning process, although they are acceptable for doing an acceptable DNN training, not are enough to structure a good dataset to provide the possibility of carrying out training tasks of a learning system of high quality. As a consequence, it is crucial to have a final dataset of high quality, to get updated data, with concrete and accuracy proprieties and attributes to conduct a better precision in prediction quality metrics.

Learning models of learning machine such as deep neural network[19], have a higher efficiency and quality in the prediction in contrast to other statistical traditional models. This quality of algorithm is clearly evident in prediction and binary or multiple classification task. Nevertheless, in relation to prediction tasks in multiple regression models perform in the same way when training dataset is relatively small. Meanwhile related to classification in all sort of conditions, deep neural network preforms better than statistical models such as linear regression.

For this reason, not all learning machine algorithms in his characterization get the same level of performance, in this case, the decision tree and deep neural networks, were algorithms those get the best results in the predictions [20, 21] according to the scientific literature review and experimental analysis for this kind of research study. In conclusion, it is important to mention that was applied a multiple layer model to overpopulate the missing data within UPTC university, sectional Sogamoso, for detecting or stand out those profiles that fit among students with a possible dropout rate.

Conclusions

The algorithm presents good results when it comes to predicting whether or not a student is at risk of dropping out, this despite not having a greater amount of information for input data for their corresponding training, the results obtained by the different techniques show high efficiency of the predictor algorithm and the confusion matrix with their metrics.

The data obtained in the preparation and cleaning process, although they are acceptable to carry out an acceptable training to the DNN, are not acceptable to structure a good data set that leads to carrying out training tasks of a high-quality learning system. It is imperative that in order to obtain a good quality final data set, more up-to-date data are obtained, with more specific and precise attributes and properties that can lead to better accuracy in prediction quality metrics. Machine learning models such as deep neural networks have higher prediction efficiency and quality than traditional statistical models.

This quality of the algorithm is clearly evidenced in binary or multiple prediction and classification tasks. Now, in relation to prediction tasks in multiple regression models, they tend to perform equally when the training data set is relatively small, while in relation to classification in all types of conditions, neural networks perform better than statistical models. like linear regression. Not all machine learning algorithms in their characterization obtain the same level of performance, in this case the decision trees and DNN were those algorithms that, according to the review of the scientific literature, are best coupled to this type of study carried out. Finally, to detect or highlight those profiles that fit among those students with a possible dropout rate, the multilayer model was applied to overpopulate the missing information within the final data set.

Bibliographic references

1. R.L. do Nascimento, R. Fagundes, & R.M, de Souza, Statistical Learning for Predicting School Dropout in Elementary Education: A Comparative Study. Annals of Data Science, vol 1,pp.1-28, 2021. https://doi.org/10.1007/s40745-021-00321-4
2. O. Spositto, M. Etcheverry, H. Ryckeboer, & J. Bossero, “Aplicación de técnicas de mineria de datos para la evaluación del rendimiento académico y la deserción estudiantil.” Novena Conferencia Iberoamericana En Sistemas, Cibernética e Informática, CISCI, 29, 2–6. DOI: https://doi.org/10.18259/acs.2018005
3. B. Castro-Montoya, C.M. Lopera-Gómez,R.D. Manrique-Hernández, & D. Gonzalez-Gómez, “A competing dropout and graduation risk survival analysis of undergraduate students at a private university in Medellín (Colombia).” *Formación universitaria*, vol *14*(1), pp. 81-9, 2021. DOI:https://doi.org/10.4067/S0718-50062021000100081
4. H. Luan, & C.Tsai. “A Review of Using Machine Learning Approaches for Precision Education”. *Educational Technology & Society*, vol *24*(1), pp. 250–266. 2021. DOI:https:// A Review of Using Machine Learning Approaches for Precision Education .www.jstor.org/stable/26977871
5. T. Cardona, E. Cudney, R. Hoerl, & J. Snyder, “Data Mining and Machine Learning Retention Models in Higher Education”. *Journal of College Student Retention: Research, Theory & Practice*, 2020. DOI:https://doi.org/10.1177/1521025120964920
6. M. Ha & H. Ahn, "A Machine Learning-Based Vocational Training Dropout Prediction Model Considering Structured and Unstructured Data," Journal of the Korea Contents Association, Vol.19, No.1, 2019,DOI: https://doi.org/10.5392/JKCA.2019.19.01.001
7. W. Cho & M. Yu, "Creating Value for Education through Big Data Analysis Education Programs", The Journal of BIGDATA, Vol.3, No.2, 2018, pp.123-130. 2020, DOI:https://doi.org/10.36498/kbigdt.2018.3.2.123
8. A. Mubarak, A. Ahmed, C.Han , & M. Ibrahim. "Deep analytic model for student dropout prediction in massive open online courses." *Computers & Electrical Engineering* , vol 93, 2021, DOI: https://doi.org/10.1016/j.compeleceng.2021.107271

9. F. Qian, et al. "CLSA: A novel deep learning model for MOOC dropout prediction." *Computers & Electrical Engineering* , vol 94, 2021. DOI: https://doi.org/10.1016/j.compeleceng.2021.107315

10. B. Shah, & S. Margil. "A Survey on Machine Learning and Deep Learning Based Approaches for Sarcasm Identification in Social Media." *Data Science and Intelligent Applications*. Springer, Singapore, vol 40, 2021, pp. 247-259. DOI: https://doi.org/10.1016/j.cosrev.2021.100395

11. D.J. Lemay, B. Clare, & D.Tenzin. "Comparison of learning analytics and educational data mining: A topic modeling approach." *Computers and Education: Artificial Intelligence* vol 2 , 2021. DOI: https://doi.org/10.1016/j.caeai.2021.100016

12. A. Sarra, L. Fontanella & S. Di Zio. "Identifying students at risk of academic failure within the educational data mining framework". *Soc Indic Res*, vol 146, 2019, pp. 41–60. DOI: https://doi.org/10.1007/s11205-018-1901-8

13. M. Sharma, SN. Khera, PB. Sharma. "Applicability of machine learning in the measurement of emotional intelligence". *Ann Data Sci*, vol 6, 2019, pp. 179–187. DOI: https://doi.org/10.1007/s40745-018-00185-1

14. C. Garbin, X. Zhu & O. Marques. "Dropout vs. batch normalization: an empirical study of their impact to deep learning". *Multimed Tools Appl,* vol 79, 2020, pp. 12777–12815. DOI: https://doi.org/10.1007/s11042-019-08453-9

15. A. Scheunemann, T. Schnettler, J. Bobe, et al. "A longitudinal analysis of the reciprocal relationship between academic procrastination, study satisfaction, and dropout intentions in higher education". *Eur J Psychol Educ*, vol 1, 2021. DOI: https://doi.org/10.1007/s10212-021-00571-z

16. A. Cohen, "Analysis of student activity in web-supported courses as a tool for predicting dropout". *Education Tech Research, vol* 65, 2017, pp 1285–1304, DOI: https://doi.org/10.1007/s11423-017-9524-3

17. J. Zhang, M. Gao, & J. Zhang, "The learning behaviours of dropouts in MOOCs: A collective attention network perspective", *Computers & Education*, vol. 167, 2021, pp. 104189, DOI: https://doi:10.1016/j.compedu.2021.104189

18. M. Youssef, S. Mohammed, E. K. Hamada, and B. F. Wafaa, "A predictive approach based on efficient feature selection and learning algorithms' competition: Case of learners' dropout in MOOCs", *Education and Information Technologies*, vol. 24, no. 6, 2019, pp. 3591–3618, DOI: https://doi:10.1007/s10639-019-09934-y

19. M. Şahin, "A Comparative Analysis of Dropout Prediction in Massive Open Online Courses", *Arabian Journal for Science and Engineering*, vol. 46, no. 2, 2019, pp. 1845–1861, DOI: https://doi:10.1007/s13369-020-05127-9

20. H. Camacho, D. Campos, I. Mercado, N. Cubillán, G. Castellar, "Uso de la cáscara de papa (Solanum tuberosum L.) en la clarificación del agua de la Ciénaga de Malambo," Investigación e Innovación en Ingenierías., vol. 8, no. 1, pp. 100–111, 2020. DOI: https://doi.org/10.17081/invinno.8.1.3572

21. N. Chen, J. Zhu, J. Chen, & T. Chen, "Dropout training for SVMs with data augmentation", *Frontiers of Computer Science*, vol. 12, no. 4, 2018, pp. 694–713, DOI: https://doi:10.1007/s11704-018-7314-7